

# Capacity Region of the Finite-State Multiple Access Channel with and without Feedback

Haim Permuter and Tsachy Weissman

## Abstract

The capacity region of the Finite-State Multiple Access Channel (FS-MAC) with feedback that may be an arbitrary time-invariant function of the channel output samples is considered. We characterize both an inner and an outer bound for this region, using Massey's directed information. These bounds are shown to coincide, and hence yield the capacity region, of FS-MACs where the state process is stationary and ergodic and not affected by the inputs. Though 'multi-letter' in general, our results yield explicit conclusions when applied to specific scenarios of interest. E.g., our results allow us to:

- Identify a large class of FS-MACs, that includes the additive mod-2 noise MAC where the noise may have memory, for which feedback does not enlarge the capacity region.
- Deduce that, for a general FS-MAC with states that are not affected by the input, if the capacity (region) without feedback is zero, then so is the capacity (region) with feedback.
- Deduce that the capacity region of a MAC that can be decomposed into a 'multiplexer' concatenated by a point-to-point channel (with, without, or with partial feedback), the capacity region is given by  $\sum_m R_m \leq C$ , where  $C$  is the capacity of the point to point channel and  $m$  indexes the encoders. Moreover, we show that for this family of channels source-channel coding separation holds.

## Index Terms

Feedback capacity, multiple access channel, capacity region, directed information, causal conditioning, code-tree, source-channel coding separation, sup-additivity of sets.

## I. INTRODUCTION

The Multiple Access Channel (MAC) has received much attention in the literature. To put our contributions in context, we begin by briefly describing some of the key results in the area. The capacity region for the memoryless MAC was derived by Ahlswede in [1]. Cover and Leung derived an achievable region for a memoryless MAC with feedback in [2]. Using block Markov encoding, superposition and list codes, they showed that the region  $R_1 \leq I(X_1; Y|X_2, U)$ ,  $R_2 \leq I(X_2; Y|X_1, U)$  and  $R_1 + R_2 \leq I(X_1, X_2; Y)$  where  $P(u, x_1, x_2, y) = p(u)p(x_1|u)p(x_2|u)p(y|x_1, x_2)$  is achievable for a memoryless MAC with feedback. Willems showed in [3] that

The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA. (Email: {haim1, tsachy}@stanford.edu)

This work was supported by the NSF through the CAREER award and TFR-0729119 grant.

the achievable region given by Cover and Leung for a memoryless channel with feedback is optimal for a class of channels where one of the inputs is a deterministic function of the output and the other input. More recently Bross and Lapidoth [4] improved Cover and Leung's region, and Wu et. al. [5] have extended Cover and Leung's region for the case that non-causal state information is available at both encoders.

Ozarow derived the capacity of a memoryless Gaussian MAC with feedback in [6], and showed it to be achievable via a modification of the Schalkwijk-Kailath scheme [7]. In general, the capacity in the presence of noisy feedback is an open question for the point-to-point channel and a fortiori for the MAC. Lapidoth and Wigger [8] presented an achievable region for the case of the Gaussian MAC with noisy feedback and showed that it converges to Ozarow's noiseless-feedback sum-rate capacity as the feedback-noise variance tends to zero. Other recent variations on the Schalkwijk-Kailath scheme of relevance to the themes of our work include the case of quantization noise in the feedback link [9] and the case of interference known non-causally at the transmitter [10].

Verdú characterized the capacity region of a Multi-Access channel of the form  $P(y_i|x_1^i, x_2^i, y^{i-1}) = P(y_i|x_{1,i-m}^i, x_{2,i-m}^i)$  without feedback in [11]. Verdú further showed in that work that in the absence of frame synchronism between the two users, i.e., there is a random shift between the users, only stationary input distributions need be considered. Cheng and Verdú built on the capacity result from [11] in [12] to show that for a Gaussian MAC there exists a water-filling solution that generalizes the point-to-point Gaussian channel.

In [13] [14], Kramer derived several capacity results for discrete memoryless networks with feedback. By using the idea of code-trees instead of code-words, Kramer derived a 'multi-letter' expression for the capacity of the discrete memoryless MAC. One of the main results we develop in the present paper extends Kramer's capacity result to the case of a stationary and ergodic Markov Finite-State MAC (FS-MAC), to be formally defined below.

In [15] [16], Han used the information-spectrum method in order to derive the capacity of a general MAC without feedback, when the channel transition probabilities are arbitrary for every  $n$  symbols. Han also considered the additive mod- $q$  MAC, which we shall use here to illustrate the way in which our general results characterize special cases of interest. In particular, our results will imply that feedback does not increase the capacity region of the additive mod- $q$  MAC.

In this work, we consider the capacity region of the Finite-State Multiple Access Channel (FS-MAC), with feedback that may be an arbitrary time-invariant function of the channel output samples. We characterize both an inner and an outer bound for this region. We further show that these bounds coincide, and hence yield the capacity region, for the important subfamily of FS-MACs with states that evolve independently of the channel inputs. Our derivation of the capacity region is rooted in the derivation of the capacity of finite-state channels in Gallager's book [17, ch 4.5]. More recently, Lapidoth and Telatar [18] have used it in order to derive the capacity of a compound channel without feedback, where the compound channel consists of a family of finite-state channels. In particular, they have introduced into Gallager's proof the idea of concatenating codewords, which we extend here to concatenating code-trees.

Though 'multi-letter' in general, our results yield explicit conclusions when applied to more specific families of MACs. For example, we find that feedback does not increase the capacity of the mod- $q$  additive noise MAC

(where  $q$  is the size of the common alphabet of the input, output and noise), regardless of the memory in the noise. This result is in sharp contrast with the finding of Gaarder and Wolf in [19] that feedback can increase the capacity even of a *memoryless* MAC due to cooperation between senders that it can create. Our result should also be considered in light of Alajaji's work [20], where it was shown that feedback does not increase the capacity of discrete point-to-point channels with mod- $q$  additive noise. Thus, this part of our contribution can be considered a multi-terminal extension of Alajaji's result. Our results will in fact allow us to identify a class of MACs larger than that of the mod- $q$  additive noise MAC for which feedback does not enlarge the capacity region.

Further specialization of the results will allow us to deduce that, for a general FS-MAC with states that are not affected by the input, if the capacity (region) without feedback is zero, then so is the capacity (region) with feedback. It will also allow us to identify a large class of FS-MACs for which source-channel coding separation holds.

The remainder of this paper is organized as follows. We concretely describe our channel model and assumptions in Section II. In Section III we introduce some notation, tools and results pertaining to directed information and the notion of causal conditioning that will be key in later sections. We state our main results in Section IV. In Section V we apply the general results of Section IV to obtain the capacity region for several interesting classes of channels, as well as establish a source-channel separation result. The validity of our inner and outer bounds is established, respectively, in Section VI and Section VII. In Section VIII we show that our inner and outer bounds coincide, and hence yield the capacity region, when applied to the FS-MAC without feedback. This result can be thought of as the natural extension of Gallager's results [17, Ch. 4] to the MAC or, alternatively, as the natural extension of Gallager's derivation of the MAC capacity region in [21] to channels with states. In Section IX we characterize the capacity region for the case of arbitrary (time-invariant) feedback and FS-MAC channels with states that evolve independently of the input, as well as the FS-MAC with limited ISI (which is the natural MAC-analogue of Kim's point-to-point channel [22]), by showing that our inner and outer bounds coincide for this case. We conclude in Section X with a summary of our contribution and a related future research direction.

## II. CHANNEL MODEL

In this paper, we consider an FS-MAC (Finite state MAC) with a time invariant feedback as illustrated in Fig. 1. The MAC setting consists of two senders and one receiver. Each sender  $l \in \{1, 2\}$  chooses an index  $m_l$  uniformly from the set  $\{1, \dots, 2^{nR_l}\}$  and independently of the other sender. The input to the channel from encoder  $l$  is denoted by  $\{X_{l1}, X_{l2}, X_{l3}, \dots\}$ , and the output of the channel is denoted by  $\{Y_1, Y_2, Y_3, \dots\}$ . The state at time  $i$ , i.e.,  $S_i \in \mathcal{S}$ , takes values in a finite set of possible states. The channel is stationary and is characterized by a conditional probability  $P(y_i, s_i | x_{1i}, x_{2i}, s_{i-1})$  that satisfies

$$P(y_i, s_i | x_1^i, x_2^i, s^{i-1}, y^{i-1}) = P(y_i, s_i | x_{1i}, x_{2i}, s_{i-1}), \quad (1)$$

where the superscripts denote sequences in the following way:  $x_l^i = (x_{l1}, x_{l2}, \dots, x_{li})$ ,  $l \in \{1, 2\}$ . We assume a communication with feedback  $z_l^i$  where the element  $z_{li}$  is a time-invariant function of the output  $y_i$ . For example,

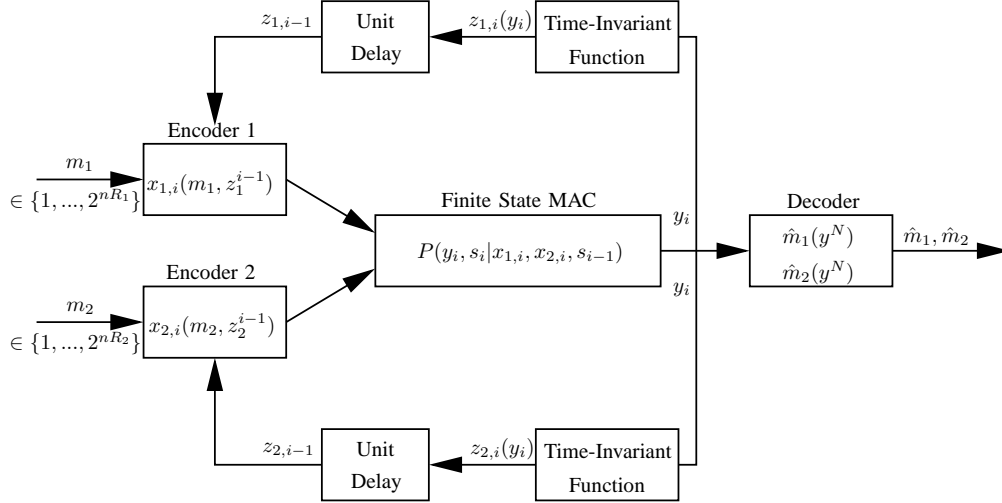


Fig. 1. Channel with feedback that is a time invariant deterministic function of the output.

$z_{li}$  could equal  $y_i$  (perfect feedback), or a quantized version of  $y_i$ , or null (no feedback). The encoders receive the feedback samples with one unit delay.

A code with feedback consists of two encoding functions  $g_l : \{1, \dots, 2^{nR_1}\} \times \mathcal{Z}_l^{n-1} \rightarrow \mathcal{X}_l^n$ ,  $l = 1, 2$ , where the  $k$ th coordinate of  $x_l^n \in \mathcal{X}_l^n$  is given by the function

$$x_{lk} = g_{lk}(m_l, z_l^{k-1}), \quad k = 1, 2, \dots, n, \quad l = 1, 2 \quad (2)$$

and a decoding function,

$$g : \mathcal{Y}^n \rightarrow \{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nR_2}\}. \quad (3)$$

The average probability of error for  $((2^{nR_1}, 2^{nR_2}, n)$  code is defined as

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{w_1, w_2} \Pr\{g(Y^n) \neq (w_1, w_2) | (w_1, w_2) \text{ sent}\}. \quad (4)$$

A rate  $(R_1, R_2)$  is said to be *achievable* for the MAC if there exists a sequence of  $((2^{nR_1}, 2^{nR_2}, n)$  codes with  $P_e^{(n)} \rightarrow 0$ . The *capacity region* of MAC is the closure of the set of achievable  $(R_1, R_2)$  rates.

### III. DIRECTED INFORMATION

Throughout this paper we use the *Causal Conditioning* notation  $(\cdot || \cdot)$ . We denote the probability mass function (pmf) of  $Y^N$  causally conditioned on  $X^{N-d}$ , for some integer  $d \geq 0$ , as  $P(y^N || x^{N-d})$  which is defined as

$$P(y^N || x^{N-d}) \triangleq \prod_{i=1}^N P(y_i | y^{i-1}, x^{i-d}), \quad (5)$$

(if  $i - d \leq 0$  then  $x^{i-d}$  is set to null). In particular, we extensively use the cases where  $d = 0, 1$ :

$$P(y^N || x^N) \triangleq \prod_{i=1}^N P(y_i | y^{i-1}, x^i) \quad (6)$$

$$Q(x^N || y^{N-1}) \triangleq \prod_{i=1}^N Q(x_i | x^{i-1}, y^{i-1}), \quad (7)$$

where the letters  $Q$  and  $P$  are both used for denoting pmfs.

Directed information  $I(X^N \rightarrow Y^N)$  was defined by Massey in [23] as

$$I(X^N \rightarrow Y^N) \triangleq \sum_{i=1}^N I(X^i; Y_i | Y^{i-1}). \quad (8)$$

It has been widely used in the characterization of capacity of point-to-point channels [22], [24]–[29], compound channels [30], network capacity [14], [31], rate distortion [32]–[34] and computational biology [35], [36]. Directed information can also be expressed in terms of causal conditioning as

$$I(X^N \rightarrow Y^N) = \sum_{i=1}^N I(X^i; Y_i | Y^{i-1}) = \mathbf{E} \left[ \log \frac{P(Y^N || X^N)}{P(Y^N)} \right], \quad (9)$$

where  $\mathbf{E}$  denotes expectation. The directed information from  $X^N$  to  $Y^N$ , conditioned on  $S$ , is denoted as  $I(X^N \rightarrow Y^N | S)$  and is defined as:

$$I(X^N \rightarrow Y^N | S) \triangleq \sum_{i=1}^N I(X^i; Y_i | Y^{i-1}, S). \quad (10)$$

Directed information between  $X_1^N$  to  $Y^N$  causally conditioned on  $X_2^N$  is defined as

$$I(X_1^N \rightarrow Y^N || X_2^N) \triangleq \sum_{i=1}^N I(X_1^i; Y_i | X_2^i, Y^{i-1}) = \mathbf{E} \left[ \log \frac{P(Y^N || X_1^N, X_2^N)}{P(Y^N || X_2^N)} \right]. \quad (11)$$

where  $P(y^N || x_1^N, x_2^N) = \prod_{i=1}^N P(y_i | y^{i-1}, x_1^i, x_2^i)$ .

Throughout this paper we are using several properties of causal conditioning and directed information that follow from the definitions and simple algebra. Many of the key properties that hold for mutual information and regular conditioning carry over to directed information and causal conditioning, where  $P(x^N)$  is replaced by  $P(x^N || y^{N-1})$  and  $P(y^N)$  is replaced by  $P(y^N || x^N)$ . Specifically,

*Lemma 1: (Analogue to  $P(x_1^N, y^N) = P(x_1^N)P(y^N | x_1^N)$ .)* For arbitrary random vectors  $(X_1^N, X_2^N, Y^N)$ ,

$$P(x_1^N, y^N) = P(x_1^N || y^{N-1})P(y^N || x_1^N) \quad (12)$$

$$P(x_1^N, y^N || x_2^N) = P(x_1^N || y^{N-1}, x_2^N)P(y^N || x_1^N, x_2^N). \quad (13)$$

*Lemma 2: (Analogue to  $|I(X_1^N; Y^N) - I(X_1^N; Y^N | S)| \leq H(S)$ .)* For arbitrary random vectors and variables,

$$|I(X_1^N \rightarrow Y^N) - I(X_1^N \rightarrow Y^N | S)| \leq H(S) \leq \log |S| \quad (14)$$

$$|I(X_1^N \rightarrow Y^N || X_2^N) - I(X_1^N \rightarrow Y^N || X_2^N, S)| \leq H(S) \leq \log |S|. \quad (15)$$

The proofs of Lemma 1 and Lemma 2 can be found in [27, Sec. IV], along with some additional properties of causal conditioning and directed information. The next lemma, which is proven in Appendix I, shows that by replacing regular pmf with causal conditioning pmf we get the directed information. Let us denote the mutual information  $I(X_1^N; Y^N | X_2^N)$  as a functional of  $Q(x_1^N, x_2^N)$  and  $P(y^N | x_1^N, x_2^N)$ , i.e.,  $\mathcal{I}(Q(x_1^N, x_2^N); P(y^N | x_1^N, x_2^N)) \triangleq$

$I(X_1^n; Y^n | X_2^n)$ . Consider the case that the random variables  $X_1^N, X_2^N$  are independent, i.e.,  $Q(x_1^N, x_2^N) = Q(x_1^N)Q(x_2^N)$ , then by definition

$$\mathcal{I}(Q(x_1^N)Q(x_2^N); P(y^N | x_1^N, x_2^N)) \triangleq \sum_{y^N, x_1^N, x_2^N} Q(x_1^N)Q(x_2^N)P(y^N | x_1^N, x_2^N) \frac{P(y^N | x_1^N, x_2^N)}{\sum_{x_1^N} Q(x_1^N)P(y^N | x_1^N, x_2^N)}. \quad (16)$$

*Lemma 3:* If the random vectors  $X_1^N$  and  $X_2^N$  are causal-conditionally independent given  $Y^{N-1}$ , i.e.,  $Q(x_1^N, x_2^N | y^{N-1}) = Q(x_1^N | y^{N-1})Q(x_2^N | y^{N-1})$  then

$$\mathcal{I}(Q(x_1^N | y^{N-1})Q(x_2^N | y^{N-1}); P(y^N | x_1^N, x_2^N)) = I(X_1^N \rightarrow Y^N | X_2^N). \quad (17)$$

The next lemma, which is proven in Appendix II, shows that in the absence of feedback, mutual information becomes directed information.

*Lemma 4:* If  $Q(x_1^N, x_2^N | y^{N-1}) = Q(x_1^N)Q(x_2^N)$  then

$$I(X_1^N; Y^N | X_2^N) = I(X_1^N \rightarrow Y^N | X_2^N). \quad (18)$$

#### IV. MAIN THEOREMS

We dedicate this section to a statement of our main results, proofs of which will appear in the subsequent sections. Let  $\underline{\mathcal{R}}_n$  denote the following region in  $\mathbb{R}_+^2$  (2D set of nonnegative real numbers):

$$\underline{\mathcal{R}}_n = \bigcup_{Q(w)Q(x_1^n | z_1^{n-1}, w)Q(x_2^n | z_2^{n-1}, w)} \begin{cases} R_1 \leq \min_{s_0} \frac{1}{n} I(X_1^n \rightarrow Y^n | X_2^n, W, s_0) - \frac{\log |\mathcal{S}|}{n}, \\ R_1 \leq \min_{s_0} \frac{1}{n} I(X_2^n \rightarrow Y^n | X_1^n, W, s_0) - \frac{\log |\mathcal{S}|}{n}, \\ R_1 + R_2 \leq \min_{s_0} \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n | W, s_0) - \frac{\log |\mathcal{S}|}{n}. \end{cases} \quad (19)$$

Having the auxiliary random variable  $W$  is equivalent to taking the convex hull of the region. It is shown in the Appendix that the inclusion (or omission) of  $W$  in the definition of the region  $\underline{\mathcal{R}}_n$  has vanishing effect with increasing  $n$ .

*Theorem 5: (Inner bound.)* For any FS-MAC with time invariant feedback as shown in Fig. 1, and for any integer  $n \geq 1$ , the region  $\underline{\mathcal{R}}_n$  is achievable.

Let  $\mathcal{R}_n$  denote the following region in  $\mathbb{R}_+^2$

$$\mathcal{R}_n = \bigcup_{Q(x_1^n | z_1^{n-1})Q(x_2^n | z_2^{n-1})} \begin{cases} R_1 \leq \frac{1}{n} I(X_1^n \rightarrow Y^n | X_2^n), \\ R_1 \leq \frac{1}{n} I(X_2^n \rightarrow Y^n | X_1^n), \\ R_1 + R_2 \leq \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n). \end{cases} \quad (20)$$

In the following theorem we use the standard notion of convergence of sets. Confer Appendix IV for the details of the definition.

*Theorem 6: (Outer bound.)* Let  $(R_1, R_2)$  be an achievable pair for a FS-MAC with time invariant feedback, as shown in Fig. 1. Then, for any  $n$  there exists a distribution  $Q(x_1^n | z_1^{n-1})Q(x_2^n | z_2^{n-1})$  such that the following

inequalities hold:

$$\begin{aligned} R_1 &\leq \frac{1}{n} I(X_1^n \rightarrow Y^n || X_2^n) + \epsilon_n \\ R_2 &\leq \frac{1}{n} I(X_2^n \rightarrow Y^n || X_1^n) + \epsilon_n \\ R_1 + R_2 &\leq \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n) + \epsilon_n, \end{aligned} \quad (21)$$

where  $\epsilon_n$  goes to zero as  $n$  goes to infinity. Moreover, the outer bound can be written as  $\liminf \mathcal{R}_n$ .

For the case where there is no feedback, i.e.,  $z_i$  is null,  $\mathcal{R}_n$  and  $\underline{\mathcal{R}}_n$  can be expressed in terms of mutual information and regular conditioning due to Lemma 4.

*Theorem 7: (Capacity of FS-MAC without feedback.)* For any indecomposable FS-MAC without feedback, the achievable region is  $\lim_{n \rightarrow \infty} \mathcal{R}_n$ , and the limit exists.

*Theorem 8: (Capacity of FS-MAC with feedback.)* For any FS-MAC of the form

$$P(y_i, s_i | x_{1i}, x_{2i}, s_{i-1}) = P(s_i | s_{i-1}) P(y_i | x_{1i}, x_{2i}, s_{i-1}), \quad (22)$$

where the state process  $S_i$  is stationary and ergodic, the achievable region is  $\lim_{n \rightarrow \infty} \mathcal{R}_n$ , and the limit exists.

The next theorems will be seen to be consequences of the capacity theorems given above.

*Theorem 9:* For the channel described in (22), where the state process  $s_i$  is stationary and ergodic, if the capacity without feedback is zero, then it is also zero in the case that there is feedback.

*Corollary 10:* For a memoryless MAC, the capacity with feedback is zero if and only if it is zero without feedback.

*Corollary 11:* Feedback does not enlarge the capacity region of a discrete additive (mod- $|\mathcal{X}|$ ) noise MAC.

In fact, among other results, we will see in the next section that the (mod- $|\mathcal{X}|$ ) noise MAC is only a subset of a larger family of MACs for which feedback does not enlarge the capacity region.

## V. APPLICATIONS

The capacity formula of a FS-MAC given in Theorems 7 and 8 is a multi-letter characterization. In general, it is very hard to evaluate it but, for the finite state point to point channel, there are several cases where the capacity with and without feedback was found numerically [37] [38], [26], [25] and analytically [28].<sup>1</sup>

The multi-letter capacity expression is also valuable for deriving useful concepts in communication. For instance, in order to show that feedback does not increase the capacity of a memoryless channel (cf. [43]), we can use the multi-letter upper bound of a channel with memory. Further, in [27] it was shown that for the cases where the capacity is given by the multi-letter expression  $C = \lim_{N \rightarrow \infty} \frac{1}{N} \max_{Q(x^N || z^{N-1})} I(X^N \rightarrow Y^N)$ , the source-channel coding separation holds. It was also shown that if the state of the channel is known at both the encoder and decoder and the channel is connected (i.e., every state can be reached with some positive probability from every other state under some input distribution), then feedback does not increase the capacity of the channel.

<sup>1</sup>For the Gaussian case without feedback there exists the water filling solution [39], and recently the feedback capacity was found analytically, for the case that the noise is an ARMA(1)-Gaussian process (cf. [40]–[42]).

In this section we use the capacity formula in order to derive three conclusions:

- 1) For a stationary ergodic Markovian channels, the capacity is zero if and only if the capacity with feedback is zero.
- 2) Identify FS-MACs that feedback does not enlarge the capacity and show that for a MAC that can be decomposed into a ‘multiplexer’ concatenated by a point-to-point channel (with, without, or with partial feedback), the capacity region is given by  $\sum_m R_m \leq C$ , where  $C$  is the capacity of the point to point channel.
- 3) Source-channel coding separation holds for a MAC that can be decomposed into a ‘multiplexer’ concatenated by a point-to-point channel (with, without, or with partial feedback).

As a special case of the second concept we show that the capacity of a Binary Gilbert-Elliot MAC is  $R_1 + R_2 \leq 1 - H(\mathcal{V})$  where  $\mathcal{V}$  is the entropy rate of the hidden Markov noise that specifies the Binary Gilbert-Elliot MAC.

#### A. Zero capacity

The first concept is given in Theorem 9 and is proved here. The proof of Theorem 9 is based on the following lemma which is proven in Appendix III.

*Lemma 12:* For a MAC described by an arbitrary causal conditioning  $p(y^n || x_1^n, x_2^n)$  the following holds:

$$\max_{Q(x_1^n || y^{n-1})Q(x_2^n || y^{n-1})} I(X_1^n, X_2^n \rightarrow Y^n) = 0 \iff \max_{Q(x_1^n)Q(x_2^n)} I(X_1^n, X_2^n \rightarrow Y^n) = 0, \quad (23)$$

and each condition also implies that  $P(y^n || x_1^n, x_2^n) = P(y^n)$  for all  $x_1^n, x_2^n$ .

*Proof of Theorem 9:* Since the channel is a Markovian channel, i.e.,

$$P(y_i, s_i | x_{1,i}, x_{2,i}, s_{i-1}) = p(s_i | s_{i-1})P(y_i | x_{1,i}, x_{2,i}, s_{i-1}) \quad (24)$$

and stationary and ergodic, its capacity region is given in Theorem 8 as  $C = \lim_{n \rightarrow \infty} \mathcal{R}_n$ . Furthermore, since the sequence  $\{\mathcal{R}_n\}$  is sup-additive (Lemma 22), then according to Lemma 23 that is given in Appendix IV  $\lim_{n \rightarrow \infty} \mathcal{R}_n = \text{cl} \left( \bigcup_{n \geq 1} \mathcal{R}_n \right)$ , implying that if the capacity without feedback is zero, then for all  $n \geq 1$

$$\max_{Q(x_1^n)Q(x_2^n)} I(X_1^n, X_2^n \rightarrow Y^n) = 0. \quad (25)$$

According to Lemma 12, the maximization of the objective in eq. (25) over the distribution  $Q(x_1^n || y^{n-1})Q(x_2^n || y^{n-1})$  is still zero, hence, the capacity region is zero even if there is perfect feedback.

■

Corollary 10, which states that the capacity of a memoryless MAC without feedback is zero if and only if the capacity with feedback is zero, follows immediately from Theorem 9 because a memoryless MAC can be considered a FS-MAC with one state.

Clearly, Theorem 9 also holds for the case of a stationary and ergodic FS-Markov point-to-point channel because a MAC is an extension of a point-to-point channel. However, it does not hold for the case of a broadcast channel. For instance, consider the binary broadcast channel given by  $y_{1,i} = x \oplus n_i$  and  $y_{2,i} = x \oplus n_{i-1}$ , where  $n_i$  is an i.i.d



Bernoulli( $\frac{1}{2}$ ) and  $\oplus$  denotes addition mod-2. The capacity without feedback is clearly zero, but if the transmitter has feedback, namely if it knows  $y_{1,i-1}$  and  $y_{2,i-1}$  at time  $i$ , then it can compute the noise  $n_{i-1} = y_{1,i-1} \oplus x_{i-1}$  and therefore it can transmit 1 bit per channel use to the second user.

*B. Examples of channels for which feedback does not enlarge capacity*

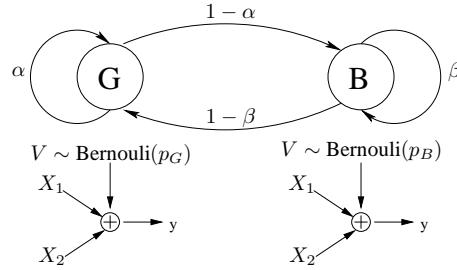


Fig. 2. Gilbert-Elliott Mac. It has two states, “Good” and “Bad” where the transition between them is according to a first order Markov process. Given that the channel is in a “Good” (or a “Bad”) state, it behaves as binary additive noise where the noise is Bernoulli( $p_G$ ) (or Bernoulli( $p_B$ ))

1) *Gilbert-Elliott MAC*: The Gilbert-Elliott channel is a widely used example of a finite state channel. It is often used to model wireless communication in the presence of fading [37], [38], [44]. The Gilbert-Elliott is a Markov channel with two states, denoted as “good” and “bad”. Each state is a binary symmetric channel and the probability of flipping the bit is lower in the “good” state. In the case of the Gillber-Elliott MAC (Fig. 2), each state is an additive MAC with i.i.d noise, where in the “good” channel the probability that the noise is ‘1’ is lower than in the bad channel. This channel can be represented as an additive MAC as in Fig. 2, where the noise is a hidden Markov process.

Since the Gilbert-Elliott MAC is an ergodic FS-MAC, its capacity with feedback when the initial state distribution over the states “good” and “bad” is the stationary distribution is given by  $\lim_{n \rightarrow \infty} \mathcal{R}_n$  (Theorem 8). For the Gilbert Elliot MAC, the region  $\lim_{n \rightarrow \infty} \mathcal{R}_n$  reduces to the simple region,

$$R_1 + R_2 \leq 1 - H(\mathcal{V}), \quad (26)$$

where  $H(\mathcal{V})$  denotes the entropy rate of the hidden Markov noise. The following equalities and inequalities upper bound the region  $\mathcal{R}_n$  and this upper bound can be achieved for any deterministic feedback by an i.i.d input distribution  $X_{1,i} \sim \text{Bernoulli}(\frac{1}{2})$  and  $X_{2,i} \sim \text{Bernoulli}(\frac{1}{2})$ ,  $i = 1, 2, \dots, n$  and  $X_1^n$  and  $X_2^n$  are independent of each other.

$$\begin{aligned}
I((X_1, X_2)^n \rightarrow Y^n) &= \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | Y^{i-1}, X_1^i, X_2^i) \\
&\stackrel{(a)}{=} \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(V_i | Y^{i-1}, X_1^i, X_2^i) \\
&= \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(V_i | V^{i-1}, Y^{i-1}, X_1^i, X_2^i) \\
&\stackrel{(b)}{=} \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(V_i | V^{i-1}) \\
&\stackrel{(c)}{\leq} \sum_{i=1}^n \log 2 - H(V_i | V^{i-1}) \\
&= n \left(1 - \frac{H(V^n)}{n}\right). \tag{27}
\end{aligned}$$

Equality (a) is due to the facts that  $y_i$  is a function of  $(v_i, x_{1,i}, x_{2,i})$  and  $v_i$  is a deterministic function of  $(y_i, x_{1,i}, x_{2,i})$ , i.e.  $y_i = x_{1,i} \oplus x_{2,i} \oplus v_i$  and  $v_i = y_i \oplus x_{1,i} \oplus x_{2,i}$ . Equality (b) follows from the fact that  $v_i$  is independent of the messages. Inequality (c) is due to the fact that the size of the alphabet  $\mathcal{Y}$  is 2. Similarly  $\frac{1}{n}I(X_1^n \rightarrow Y^n | X_2^n) \leq 1 - \frac{H(V^n)}{n}$ , and  $\frac{1}{n}I(X_2^n \rightarrow Y^n | X_1^n) \leq 1 - \frac{H(V^n)}{n}$  and equality is achieved with an i.i.d input distribution Bernoulli( $\frac{1}{2}$ ). Finally, by dividing both sides by  $n$  and using the definition of entropy rate  $H(\mathcal{V}) = \lim_{n \rightarrow \infty} \frac{1}{n}H(V^n)$  we conclude the proof.

2) *Multiplexer followed by a point-to-point channel:* Here we extend the Gilber-Elliot MAC to the case where the discrete MAC can be decomposed into two components as shown in Fig. 3. The first component is a MAC that can behave as a multiplexer and the second component is a point-to-point channel. The definitions of those components are the following:

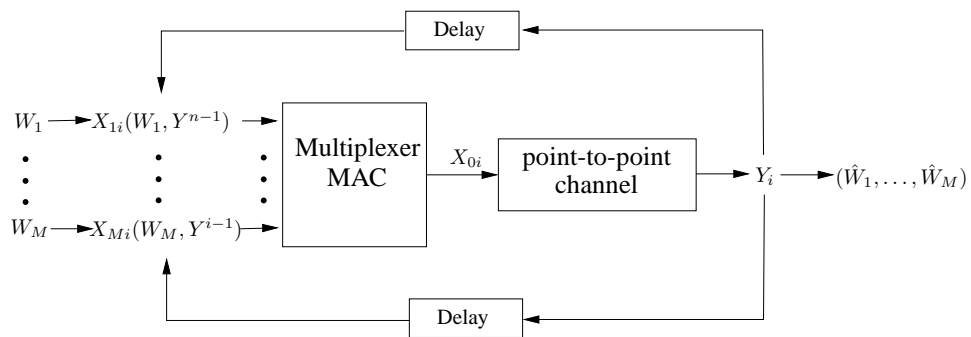


Fig. 3. Discrete MAC that can be decomposed into two parts. The first part is a MAC that behaves as a multiplexer and the second part is a point-to-point channel

*Definition 1:* A MAC behaves as a *multiplexer* if the inputs and the output have common alphabets and for all  $m \in 1, \dots, M$  there exists a choice of input symbols for all senders except sender  $m$ , such that the output is the  $m$ th input, i.e.  $Y = X_m$ .

An example of a multiplexer-MAC for the Binary case is a MAC whose output is one of and/or/xor of the inputs. For a general alphabet  $q$  those operations could be max/min/addition-mod- $q$ . For instance, if the channel is binary with two users and it is addition-mod-2, i.e.,  $y = x_1 \oplus x_2$ , then we can ensure that  $y = x_1$  by choosing  $x_2 = 0$ .

*Theorem 13:* The capacity region of a multiplexer MAC followed by a point-to-point channel with a time invariant feedback to all encoders, as shown in Fig. 3, is

$$\sum_{m=1}^M R_m \leq C \quad (28)$$

where  $C$  is the capacity of the point-to-point channel with the time invariant feedback  $z_{i-1}(y_{i-1})$ .

*Proof:* The achievability is proved simply by time sharing. At each time, only one selected user sends information and the other users send a constant input that insures that the output is the input of the selected user.

The converse is based on the fact that the maximum rate that can be transmitted through the point-to-point channel is  $C$  and it is an upper bound sum-rate of multiplexer-MAC. If it hadn't been an upper bound for the multiplexer-MAC, we could build a fictitious Multiplexer-MAC before the point-to-point channel and achieve by that a higher rate than its upper bound which would be contradiction. ■

3) *Discrete additive MAC:* An immediate consequence of Theorem 13 is an extension of Alajaj's result [20] to the additive MAC which is given in Corollary 11. Corollary 11 states that feedback does not enlarge the capacity region of a discrete additive (mod- $|\mathcal{X}|$ ) noise MAC.

The proof of the corollary is based on the following observation. If feedback does not increase the capacity of a particular point-to-point channel then feedback also does not increase the capacity of the MUX followed by the same particular channel. Specifically, feedback does not increase the achievable region of an additive MAC (Fig. 4) and the achievable region is given by

$$\sum_{m=1}^M R_m \leq \log q - H(\mathcal{V}), \quad (29)$$

where  $H(\mathcal{V})$  is the entropy rate of the additive noise.

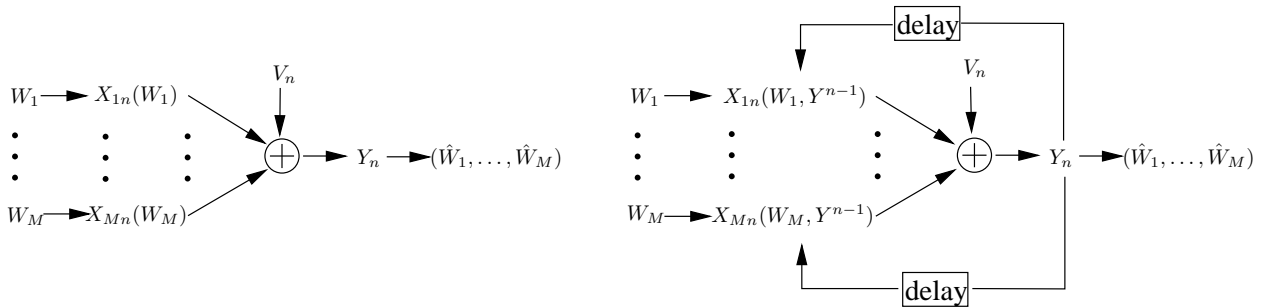


Fig. 4. Additive noise MAC with and without feedback. The random variables  $X_{1n}, \dots, X_{Mn}, Y_n, V_n, n \in 1, 2, 3, \dots$ , are from a common alphabet of size  $q$ , and they denote the input from sender 1,...,M, the output and the noise at time  $n$ , respectively. The relation between the random variables is given by  $y_n = x_{1n} \oplus x_{2n} \dots \oplus x_{Mn} \oplus v_n$  where  $\oplus$  denotes addition mod- $q$ . The noise  $V_n$ , possibly with memory, is independent of the messages  $W_1, \dots, W_M$ .

4) *Multiplexer followed by erasure channel*: Consider the case of the multiplexer-erasure MAC which is a multiplexer followed by an erasure channel, possibly with memory.

*Definition 2*: A point-to-point channel is called *erasure* channel if the output at time  $n$  can be written as  $Y_n = f(X_n, Z_n)$ , and the following properties hold:

- 1) The alphabet of  $Z$  is binary and the alphabet of  $Y$  is the same as  $X$  plus one additional symbol called the erasure.
- 2) The process  $Z_n$  is stationary and ergodic and is independent of the message.
- 3) If  $z_n = 0$ , then  $y_n = x_n$  and if  $z_n = 1$ , then the output is an erasure regardless of the input.

For the multiplexr-erasure channel we have the following theorem.

*Corollary 14*: The capacity region of the multiplexer-erasure MAC with or without feedback is

$$\sum_{m=1}^M R_m \leq (1 - p_e) \log q, \quad (30)$$

where  $p_e$  is the marginal probability of having an erasure. Moreover, even if the encoder has non causal side information, i.e. the encoders know where the erasures appear noncausally, the capacity is still given by (30).

*Proof*: According to Theorem 13 the capacity region is

$$\sum_{m=1}^M R_m \leq C, \quad (31)$$

where  $C$  is the capacity of the erasure point-to-point channel. Diggavi and Grossglauser [45, Thm. 3.1] showed that the capacity of a point-to-point erasure channel, with and without feedback, is given by  $(1 - p_e) \log q$ . Since the probability of having an erasure does not depend on the input to the channel, we deduce that even in the case where the encoder knows the sequence  $Z^n$  non-causally, which is better than feedback, the transmitter can transmit only fraction  $1 - p_e$  of the time, hence the capacity cannot exceed  $(1 - p_e) \log q$ . ■

5) *Multiplexer followed by the trapdoor channel*: In this example feedback increases the capacity. Based on the fact that the capacity of the trapdoor channel with feedback [28] is the logarithm of the golden ratio, i.e.  $\log \frac{\sqrt{5}+1}{2}$ , the achievable region of a Multiplexer followed by the trapdoor channel is

$$\sum_{m=1}^M R_m \leq \log \frac{\sqrt{5}+1}{2}. \quad (32)$$

### C. Source-channel coding separation

Cover, El-Gamal and Salehi [46] showed that, in general, the source channel separation does not hold for MACs even for a memoryless channel without feedback. However, for the case where the MAC is a discrete Multiplexer followed by a channel we now show that it does hold.

We want to send the sequence of symbols  $U_1^n, U_2^n$  over the MAC, so that the receiver can reconstruct the sequence. To do this we can use a joint source-channel coding scheme where we send through the channel the symbols  $x_{1,i}(u_1^n, z^{i-1})$  and  $x_{2,i}(u_2^n, z^{i-1})$ . The receiver looks at his received sequence  $Y^n$  and makes an estimate  $\hat{U}_1^n, \hat{U}_2^n$ . The receiver makes an error if  $\hat{U}_1^n \neq U_1^n$  or if  $\hat{U}_2^n \neq U_2^n$ , i.e., the probability of error  $P_e^{(n)}$  is  $P_e^{(n)} = \Pr((\hat{U}_1^n, \hat{U}_2^n) \neq (U_1^n, U_2^n))$ .

*Theorem 15: (Source-channel coding theorem for a Multiplexer followed by a channel.)* Let  $(U_1, U_2)_{n \geq 1}$  be a finite alphabet, jointly stationary and ergodic pair of processes and let the MAC channel be a multiplexer followed by a point-to-point channel with time invariant feedback and capacity  $C = \lim_{N \rightarrow \infty} \frac{1}{N} \max_{Q(x^n || z^{n-1})} I(X^n; Y^n)$  (e.g., a memoryless channel, an indecomposable FSC without feedback, stationary and ergodic Markovian channel). For the source and the MAC described above:

(direct part.) There exists a source-channel code with  $P_e^{(n)} \rightarrow 0$ , if  $H(\mathcal{U}_1, \mathcal{U}_2) < C$ , where  $H(\mathcal{U}_1, \mathcal{U}_2)$  is the entropy rate of the sources and  $C$  is the capacity of the point-to-point channel with a time-invariant feedback.

(converse part.) If  $H(\mathcal{U}_1, \mathcal{U}_2) > C$ , then the probability of error is bounded away from zero (independent of the blocklength).

*Proof:* The achievability is a straightforward consequence of the Slepian-Wolf result for Ergodic and stationary processes [47] and the achievability of the multiplexer followed by a point-to-point channel. First, we encode the sources by using the Sepian-Wolf achievability scheme where we assign every  $u_1^n$  to one of  $2^{nR_1}$  bins according to a uniform distribution on  $\{1, \dots, 2^{nR_1}\}$  and independently we assign every  $u_2^n$  to one of  $2^{nR_2}$  bins according to a uniform distribution on  $\{1, \dots, 2^{nR_2}\}$ . Second, we encode the bins as if they were messages, as shown in Fig. 5.

In the converse, we assume that there exists a sequence of codes with  $P_e^{(n)} \rightarrow 0$ , and we show that it implies that  $H(\mathcal{U}_1, \mathcal{U}_2) \leq C$ . Fix a given coding scheme and consider the following:

$$\begin{aligned}
H(U_1^n, U_2^n) &\stackrel{(a)}{\leq} I(U_1^n, U_2^n; \hat{U}_1^n, \hat{U}_2^n) + n\epsilon_n \\
&\stackrel{(b)}{\leq} I(U_1^n, U_2^n; Y^n) + n\epsilon_n \\
&= H(Y^n) - H(Y^n | U_1^n, U_2^n) + n\epsilon_n \\
&= \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | U_1^n, U_2^n, Y^{i-1}) + n\epsilon_n \\
&\stackrel{(c)}{=} \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | U_1^n, U_2^n, Y^{i-1}, X_1^i, X_2^i) + n\epsilon_n \\
&\stackrel{(d)}{=} \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | Y^{i-1}, X_1^i, X_2^i) + n\epsilon_n \\
&= \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | Y^{i-1}, X_1^i, X_2^i) + n\epsilon_n \\
&= \sum_{i=1}^n I(X_1^i, X_2^i; Y_i | Y^{i-1}) + n\epsilon_n \\
&\stackrel{(e)}{\leq} \sum_{i=1}^n I(X_0^i; Y_i | Y^{i-1}) + n\epsilon_n \\
&= I(X_0^n \rightarrow Y^n) + n\epsilon_n \\
&\leq \max_{Q(x_0^n || z^{n-1})} I(X_0^n \rightarrow Y^n) + n\epsilon_n
\end{aligned} \tag{33}$$

Inequality (a) is due to Fano's inequality where  $n\epsilon_n = 1 + P_e^{(n)} n |\mathcal{U}_1| |\mathcal{U}_2|$ . Inequality (b) follows from the data

processing inequality because  $(U_1^N, U_2^N) - Y^N - (\hat{U}_1^N, \hat{U}_2^N)$  form a Markov chain. Equality (c) is due to the fact that, for a given code,  $X_1^i$  is a deterministic function of  $U_1^n, Y^{i-1}$  and, similarly,  $X_2^i$  is a deterministic function of  $U_2^n, Y^{i-1}$ . Equality (d) is due to the Markov chain  $(U_1^N, U_2^N) - (X_1^i, X_2^i, Y^{i-1}) - Y_i$ . The notation  $X_{0,i}$  denotes the output of the multiplexer which is also the input to the point-to-point channel at time  $i$ . The inequality in (e) is due to the data processing inequality which can be invoked thank to the fact that given  $Y^{i-1}$  we have the Markov chain  $X_1^i, X_2^i - X_{0,i} - Y_i$ .

By dividing both sides of (33) by  $n$ , taking the limit  $n \rightarrow \infty$ , and recalling that  $C = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{Q(x^n || z^{n-1})} I(X^n; Y^n)$  we have

$$H(\mathcal{U}_1, \mathcal{U}_2) = \lim_{n \rightarrow \infty} \frac{1}{n} H(U_1^n, U_2^n) \leq C. \quad (34)$$

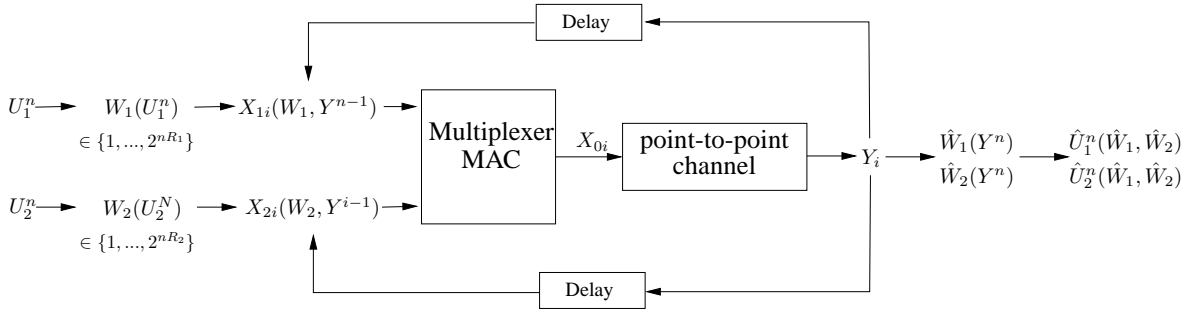


Fig. 5. Source-channel coding separation in a discrete Multiplexer followed by a point-to-point channel.

## VI. PROOF OF ACHIEVABILITY (THEOREM 5)

The proof of achievability for the FS-MAC with feedback is similar to the proof of achievability for the point-to-point FSC given in [27, Sec. V], but there are two main differences:

- 1) In the case of FSC, only one message is sent, and in the case of FS-MAC, two independent messages are sent, which requires that we analyze three different types of errors: the first type occurs when only the first message is decoded with error, the second type occurs when only the second message is decoded with error, and the third type occurs when both messages are decoded with error.
- 2) In both cases, we generate the encoding scheme (code-trees) randomly but the distribution that is used is different. In the case of FSC we generate, for each message in  $[1, \dots, 2^{NR}]$ , a code-tree of length  $N$  by using the causal conditioning distribution  $Q^*(x^N || z^{N-1}) = \arg \max_{Q(x^N || z^{N-1})} \min_{s_0} I(X^N \rightarrow Y^N | s_0)$ , and here we generate for each message in  $[1, \dots, 2^{NR_l}], l = 1, 2$  a code-tree of length  $N = Kn$  by concatenating  $K$  independent code-trees where each one is created with a causal conditioning distribution  $Q(x_l^n || z_l^{n-1}), l = 1, 2$ .

**Encoding scheme:** Randomly generate for encoder  $\{l \in 1, 2\}$ ,  $2^{NR_l}$  code-trees of length  $N = Kn$  by drawing it with the fixed distributions  $Q(x_l^n || z_l^{n-1})$ . In other words, given a feedback sequence  $z_1^{N-1}$  the causal conditioning

probability that the sequence  $x_1^N$  will be mapped to a given message is

$$Q(x_1^N | z_1^{N-1}) = \prod_{k=1}^K Q(x_{1,(k-1)n+1}^{kn} | z_{1,(k-1)n+1}^{kn}), \quad (35)$$

where  $x_{1,(k-1)n+1}^{kn}$  denotes the vector  $(x_{1,(k-1)n+1}, x_{1,(k-1)n+2}, \dots, x_{1,kn})$ . Fig. 6 illustrates the concatenation of trees graphically. In order to shorten the notation we will sometimes use the notation  $Q_N$  to denote  $Q(x_1^N | z_1^{N-1})Q(x_2^N | z_2^{N-1})$  and we will express the concatenation of pmfs in (35) as  $Q_N = \prod_{k=1}^K Q_n$ .

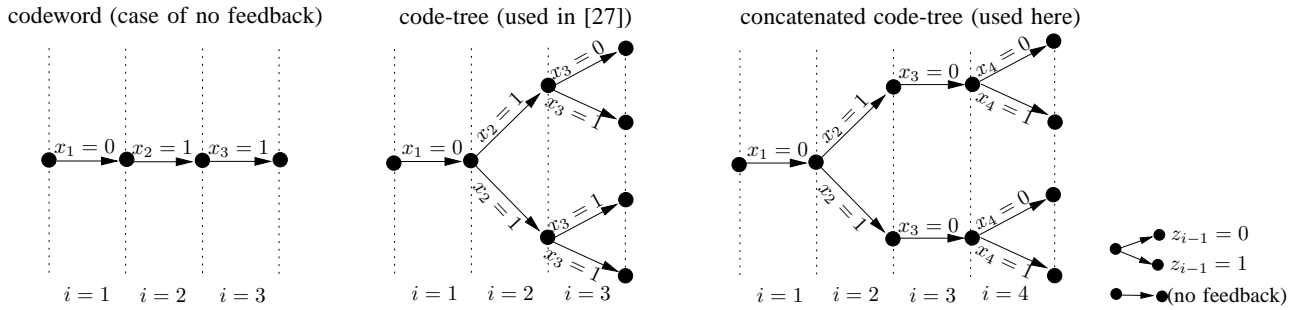


Fig. 6. Illustration of coding scheme for setting without feedback, setting with feedback as used for point-to-point channel [27] and a code-tree that was created by concatenating smaller code-trees. In the case of no feedback each message is mapped to a codeword, and in the case of feedback each message is mapped to a code-tree. The third scheme is a code-tree of depth 4 created by concatenating two trees of depth 2.

**Decoding Errors:** For each code in the ensemble, the decoder uses maximum likelihood decoding and we want to upper bound the expected value  $\mathbb{E}[P_e]$  for this ensemble. Let  $P_{e1}, P_{e2}, P_{e3}$  be defined as follows.

$P_{e1}$  (type 1 error): probability that the decoded pair  $(m_1, m_2)$  satisfies  $\hat{m}_1 \neq m_1, \hat{m}_2 = m_2$ ,

$P_{e2}$  (type 2 error): probability that the decoded pair  $(m_1, m_2)$  satisfies  $\hat{m}_1 = m_1, \hat{m}_2 \neq m_2$ ,

$P_{e3}$  (type 3 error): probability that the decoded pair  $(m_1, m_2)$  satisfies  $\hat{m}_1 \neq m_1, \hat{m}_2 \neq m_2$ .

Because the error events are disjoint we have

$$P_e = P_{e1} + P_{e2} + P_{e3} \quad (36)$$

In the next sequence of theorems and lemmas, we upper bound the expected value of each error type and show that if  $(R_1, R_2)$  satisfies the three inequalities that define  $\underline{\mathcal{R}}_n$  then the corresponding  $\mathbb{E}[P_{ei}], i = 1, 2, 3$  goes to zero and hence  $\mathbb{E}[P_e]$  goes to zero.

*Theorem 16:* Suppose that an arbitrary message  $m_1, m_2, 1 \leq m_1 \leq M_1, 1 \leq m_2 \leq M_2$ , enters the encoder with feedback and that ML decoding is employed. Let  $E[P_{e1} | m_1, m_2]$  denote the probability of decoding error averaged

over the ensemble of codes when the messages  $m_1, m_2$  were sent. Then for any choice of  $\rho, 0 < \rho \leq 1$ ,

$$\mathbf{E}[P_{e1}|m_1, m_2] \leq (M_1 - 1)^\rho \sum_{y^N, x_2^N} Q(x_2^N || z^{N-1}) \left[ \sum_{x_1^N} Q(x_1^N || z_1^{N-1}) P(y^N || x_1^N, x_2^N)^{\frac{1}{1+\rho}} \right]^{1+\rho}, \quad (37)$$

$$\mathbf{E}[P_{e2}|m_1, m_2] \leq (M_2 - 1)^\rho \sum_{y^N, x_1^N} Q(x_1^N || z_1^{N-1}) \left[ \sum_{x_2^N} Q(x_2^N || z^{N-1}) P(y^N || x_1^N, x_2^N)^{\frac{1}{1+\rho}} \right]^{1+\rho}, \quad (38)$$

$$\mathbf{E}[P_{e3}|m_1, m_2] \leq ((M_2 - 1)(M_2 - 1))^\rho \sum_{y^N} \left[ \sum_{x_1^N, x_2^N} Q(x_1^N || z_1^{N-1}) Q(x_2^N || z^{N-1}) P(y^N || x_1^N, x_2^N)^{\frac{1}{1+\rho}} \right]^{1+\rho}. \quad (39)$$

The proof is given in Appendix VI and is similar to [27, Theorem 9] only that here we take into account the fact that there are two encoders rather than one.

Let  $P_{ei}(s_0), i = 1, 2, 3$  be the probability of error of type  $i$  given that the initial state of the channel is  $s_0$ . Also let  $R_1 = \frac{1}{N} \log M_1$  and  $R_2 = \frac{1}{N} \log M_2$  be the rate of the code and  $R_3$  be the sum rate, i.e.  $R_3 = R_1 + R_2$ . The following theorem establishes exponential bounds on  $\mathbf{E}[P_{ei}(s_0)]$ .

*Theorem 17:* The average probability of error over the ensemble, for all initial states  $s_0$ , and all  $\rho, 0 \leq \rho \leq 1$ , is bounded as

$$\mathbf{E}[P_{ei}(s_0)|m_1, m_2] \leq |\mathcal{S}| 2^{\{-N[-\rho R_i + F_{N,i}(\rho, Q_N)]\}}, \quad i = 1, 2, 3 \quad (40)$$

where

$$F_{N,i}(\rho, Q_N) = -\frac{\rho \log |\mathcal{S}|}{N} + \left[ \min_{s_0} E_{N,i}(\rho, Q_N, s_0) \right], \quad i = 1, 2, 3 \quad (41)$$

$$E_{N,1}(\rho, Q_N, s_0) = -\frac{1}{N} \log \sum_{y^N, x_2^N} Q(x_2^N || z^{N-1}) \left[ \sum_{x_1^N} Q(x_1^N || z_1^{N-1}) P(y^N || x_1^N, x_2^N, s_0)^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad (42)$$

$$E_{N,2}(\rho, Q_N, s_0) = -\frac{1}{N} \log \sum_{y^N, x_1^N} Q(x_1^N || z_1^{N-1}) \left[ \sum_{x_2^N} Q(x_2^N || z^{N-1}) P(y^N || x_1^N, x_2^N, s_0)^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad (43)$$

$$E_{N,3}(\rho, Q_N, s_0) = -\frac{1}{N} \log \sum_{y^N} \left[ \sum_{x_1^N, x_2^N} Q(x_1^N || z_1^{N-1}) Q(x_2^N || z^{N-1}) P(y^N || x_1^N, x_2^N, s_0)^{\frac{1}{1+\rho}} \right]^{1+\rho}. \quad (44)$$

The proof is based on algebraic manipulation of the bounds given in (37)-(39). It is similar to the proof of Theorem 9 in [27] and therefore omitted. There are two differences between the proofs (and both are straightforward to accommodate): Here the input distribution  $Q_N = Q(x_1^N || z_1^N) Q(x_2^N || z_2^N)$  is arbitrary while in [27] we chose the one that maximizes the error exponent. Second, here we bound the averaged error over the ensemble and in [27] we have an additional step where we claim that there exists a code that has an error that is bounded by the expression in (40). Because of this difference the bound on the probability of error in [27] has an additional factor of 4.

The following theorem presents a few properties of the functions  $E_{N,i}(\rho, Q_N, s_0)$ ,  $i = 1, 2, 3$ , such as positivity of the function and its derivative, convexity with respect to  $\rho$ , and an upper bound on the derivative which is achieved for  $\rho = 0$ .



*Lemma 18:* The term  $E_{N,i}(\rho, Q_N, s_0)$  has the following properties:

$$E_{N,i}(\rho, Q_N, s_0) \geq 0; \quad \rho \geq 0, i = 1, 2, 3, \quad (45)$$

$$\begin{aligned} \frac{1}{N} I(X_1^N \rightarrow Y^N | X_2^N, s_0) &\geq \frac{\partial E_{N,1}(\rho, Q_N, s_0)}{\partial \rho} > 0; \quad \rho \geq 0 \\ \frac{1}{N} I(X_2^N \rightarrow Y^N | X_1^N, s_0) &\geq \frac{\partial E_{N,2}(\rho, Q_N, s_0)}{\partial \rho} > 0; \quad \rho \geq 0 \\ \frac{1}{N} I(X_1^N, X_2^N \rightarrow Y^N | s_0) &\geq \frac{\partial E_{N,3}(\rho, Q_N, s_0)}{\partial \rho} > 0; \quad \rho \geq 0 \end{aligned} \quad (46)$$

$$\frac{\partial^2 E_{N,i}(\rho, Q_N, s_0)}{\partial \rho^2} > 0; \quad \rho \geq 0, i = 1, 2, 3. \quad (47)$$

Furthermore, equality holds in (45) when  $\rho = 0$ , and equality holds on the left sides of eq. (46) when  $\rho = 0$  for  $i = 1, 2, 3$ .

The proof of the theorem is the same proof as [21, eq. (2.20)], [17, Theorem 5.6.3]. In [21] the arguments  $Q_N$  of  $E_{N,1}(\rho, Q_N, s_0)$  are regular conditioning i.e.,  $Q(x_1^N)Q(x_2^N)$ , and the channel is given by  $P(y^N | x_1^N, x_2^N, s_0)$ , hence the derivative of  $E_{N,1}(\rho, Q_N, s_0)$  with respect to  $\rho$  is upper-bounded by  $I(X_1^N; Y^N | X_2^N, s_0)$ . Here we replace  $Q(x_1^N)Q(x_2^N)$  with  $Q(x_1^N | z_1^{N-1})Q(x_2^N | z_2^{N-1})$  and  $P(y^N | x_1^N, x_2^N, s_0)$  with  $P(y^N | x_1^N, x_2^N, s_0)$  and, according to Lemma 3, the upper-bound becomes  $I(X_1^N \rightarrow Y^N | X_2^N, s_0)$ . The next lemma establishes the sup-additivity of  $F_{N,i}(\rho, Q_N)$ ,  $i = 1, 2, 3$ .

*Lemma 19: Sup-additivity of  $F_{N,i}(\rho, Q_N)$ .* For any finite-state channel,  $F_{N,i}(\rho, Q_N)$ , as given by eq. (41), satisfies

$$F_{n+l,i}(\rho, Q_{n+l}) \geq \frac{n}{n+l} F_{n,i}(\rho, Q_n) + \frac{l}{n+l} F_{l,i}(\rho, Q_l), \quad i = 1, 2, 3. \quad (48)$$

The proof steps are identical to the proof of the sub-additivity for the point-to-point channel [27, Lemma 11].

Invoking this lemma on the pmf  $Q_N = \prod_{k=1}^K Q_n$  where  $N = nK$  we get

$$F_{N,i}(\rho, Q_N) \geq K \frac{n}{N} F_{n,i}(\rho, Q_n) = F_{n,i}(\rho, Q_n). \quad (49)$$

Let us define

$$\underline{C}_{N,1}(Q_N) = \frac{1}{N} \min_{s_0} I(X_1^N \rightarrow Y^N | X_2^N, s_0) \quad (50)$$

$$\underline{C}_{N,2}(Q_N) = \frac{1}{N} \min_{s_0} I(X_2^N \rightarrow Y^N | X_1^N, s_0) \quad (51)$$

$$\underline{C}_{N,3}(Q_N) = \frac{1}{N} \min_{s_0} I(X_1^N, X_2^N \rightarrow Y^N | s_0) \quad (52)$$

where the joint distribution of  $X_1^N, X_2^N, Y^N$  conditioned on  $s_0$  is given by  $P(x_1^N, x_2^N, y^N | s_0) = Q(x_1^N | z_1^{N-1})Q(x_2^N | z_2^{N-1})P(y^N | x_1^N, x_2^N, s_0)$ .

Theorem 5 (inner bound) given in Sec. IV states that for every  $n$  and  $0 \leq R_i < \underline{C}_{n,i}(Q_n) - \frac{\log |\mathcal{S}|}{n}$ ,  $i = 1, 2, 3$  (recall,  $R_3 \triangleq R_1 + R_2$ ) and every  $\eta > 0$  there exists an  $N$  and an  $(N, \lceil 2^{NR_1} \rceil, \lceil 2^{NR_1} \rceil)$  code with a probability of error  $P_e(s_0)$  (averaged over the messages) that is less than  $\eta$  for all initial states  $s_0$ .

*Proof of Theorem 5:* The proof consists of the following three steps:

- Showing that for a fixed  $n$  if  $R_i < \underline{C}_{n,i}(Q_n) - \frac{\log |\mathcal{S}|}{n}$ ,  $i = 1, 2, 3$  then there exists  $\rho^*$  such that,

$$F_{n,i}(\rho^*, Q_n) - \rho^* R_i > 0, \quad i = 1, 2, 3. \quad (53)$$

- We choose  $\epsilon < \min_{i \in \{1,2,3\}} F_{n,i}(\rho^*, Q_n) - \rho^* R_i$  and show that for sufficiently large  $N$

$$\mathbf{E}[P_{ei}(s_0)|m_1, m_2] \leq 2^{-N([F_{n,i}(\rho^*, Q_n) - \rho^* R_i] - \epsilon)}, \quad \forall s_0. \quad (54)$$

- From the last step we deduce the existence of a  $(N, \lceil 2^{NR_1} \rceil, \lceil 2^{NR_1} \rceil)$  code s.t.

$$P_e(s_0) < \eta, \quad \forall s_0. \quad (55)$$

First step: for any pair  $(R_1, R_2)$ , we can rewrite eq. (40) for  $i=1,2,3$  as

$$\mathbf{E}[P_{ei}(s_0)|m_1, m_2] \leq 2^{-N(F_{N,i}(\rho, Q_N) - \rho R_i - \frac{\log |\mathcal{S}|}{N})}. \quad (56)$$

By using (49), which states that  $F_{N,i}(\rho, Q_N) \geq F_{n,i}(\rho, Q_n)$ , we get

$$\mathbf{E}[P_{ei}(s_0)|m_1, m_2] \leq 2^{-N(F_{n,i}(\rho, Q_n) - \rho R_i - \frac{\log |\mathcal{S}|}{N})}. \quad (57)$$

Note that  $F_{n,i}(\rho, Q_n)$  and therefore  $F_{n,i}(\rho, Q_n) - \rho R_i$  is continuous in  $\rho \in [0, 1]$ , so there exists a maximizing  $\rho$ . Let us show that if  $R_1 < \underline{C}_{n,1}(Q_n) - \frac{\log |\mathcal{S}|}{n}$ , then  $\max_{0 \leq \rho \leq 1} [F_{n,1}(\rho, Q_n) - \rho R_1] > 0$  (the cases  $i = 2, 3$  are identical to  $i = 1$ ). Let us define  $\delta \triangleq \underline{C}_{n,1} - R_1$ . From Lemma 18, we have that  $E_{n,1}(\rho, Q_N, s_0)$  is zero when  $\rho = 0$ , is a continuous function of  $\rho$ , and its derivative at zero with respect to  $\rho$  is equal or greater to  $\underline{C}_{n,1}$ , which satisfies  $\underline{C}_{n,1} \geq R_1 + \frac{\log |\mathcal{S}|}{n} + \frac{\delta}{2}$ . Thus, for each state  $s_0$  there is a range  $\rho > 0$  such that

$$E_{n,1}(\rho, Q_N, s_0) - \rho(R_1 + \frac{\log |\mathcal{S}|}{n}) > 0. \quad (58)$$

Moreover, because the number of states is finite, there exists a  $\rho^* > 0$  for which the inequality (58) is true for all  $s_0$ . Thus, from the definition of  $F_{n,1}(\rho^*, Q_n)$  given in (41) and from (58),

$$F_{n,1}(\rho^*, Q_n) = -\rho^* \frac{\log |\mathcal{S}|}{n} + \min_{s_0} E_{n,1}(\rho^*, Q_n, s_0) > \rho^* R_1, \quad \forall s_0. \quad (59)$$

Second step: We choose a positive number  $\epsilon$  such that  $\epsilon < \min_{i \in \{1,2,3\}} F_{n,i}(\rho^*, Q_n) - \rho^* R_i$ . It follows from (57) that for every  $N$  that satisfies  $N > \frac{\log |\mathcal{S}|}{\epsilon}$ ,

$$\mathbf{E}[P_{ei}(s_0)|m_1, m_2] \leq 2^{-N(F_{n,i}(\rho^*, Q_n) - \rho^* R_i - \epsilon)}, \quad (60)$$

and according to the first step of the proof the exponent  $F_{n,i}(\rho^*, Q_n, s_0) - \rho^* R_i - \epsilon$  is strictly positive.

Third step: According to the previous step, for all  $\frac{\eta}{3|\mathcal{S}|+1} > 0$  there exists an  $N$  such that  $\mathbf{E}[P_{ei}(s_0)|m_1, m_2] \leq \frac{\eta}{3|\mathcal{S}|+1}$  for all  $i \in 1, 2, 3$  all  $s_0 \in \mathcal{S}$  and all messages. Since  $P_e(s_0) = \sum_{i=1}^3 P_{ei}(s_0)$ , then  $\mathbf{E}[P_e(s_0)|m_1, m_2] \leq \frac{\eta}{|\mathcal{S}|+1}$ ; furthermore  $\mathbf{E}[P_e(s_0)] \leq \frac{\eta}{|\mathcal{S}|+1}$  for all  $s_0 \in \mathcal{S}$ . By using the Markov inequality, we have

$$\Pr(P_e(s_0) \geq \eta) \leq \frac{1}{|\mathcal{S}|+1}, \quad (61)$$

and by using the union bound we have

$$\Pr(P_e(s_0) \geq \eta, \text{ for some } s_0 \in \mathcal{S}) \leq \sum_{s_0 \in \mathcal{S}} \Pr(P_e(s_0) \geq \eta) = \frac{|\mathcal{S}|}{|\mathcal{S}|+1} < 1. \quad (62)$$

Because the probability over the ensemble of codes of having a code with error probability (averaged over all messages) that is less than  $\eta$  for all initial states is positive, there must exist at least one code that has an error probability (averaged over all messages) that is less than  $\eta$  for all initial states. ■

## VII. PROOF OF THE OUTER BOUND (THEOREM 6)

In this section we prove Theorem 6, which states that for any FS-MAC there exists a distribution  $Q(x_1^n || z_1^{n-1})Q(x_2^n || z_2^{n-1})$  such that the following inequalities hold:

$$\begin{aligned} R_1 &\leq \frac{1}{n} I(X_1^n \rightarrow Y^n || X_2^n) + \epsilon_n \\ R_2 &\leq \frac{1}{n} I(X_2^n \rightarrow Y^n || X_1^n) + \epsilon_n \\ R_1 + R_2 &\leq \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n) + \epsilon_n, \end{aligned} \quad (63)$$

where  $\epsilon_n$  goes to zero as  $n$  goes to infinity.

*Proof of Theorem 6:* Let  $W_1$  and  $W_2$  be two independent messages, chosen independently and according to a uniform distribution  $\Pr(W_l = w_l) = 2^{-nR_l}, l = 1, 2$ . The input to the channel from encoder  $l$  at time  $i$  is  $x_{li}$ , and is a function of the message  $W_i$  and the arbitrary deterministic feedback output  $z_l^{i-1}(y^{i-1})$ .

The following sequence of equalities and inequalities proves that if a code that achieves rate  $R_1$  exists then the

first inequality holds, i.e.,  $R_1 \leq \frac{1}{n}I(X_1^n \rightarrow Y^n | X_2^n) + \epsilon_n$ :

$$\begin{aligned}
nR_1 &\stackrel{(a)}{=} H(W_1) \\
&\stackrel{(b)}{=} H(W_1 | W_2) \\
&= I(W_1; Y^n | W_2) + H(W_1 | Y^n, W_2) \\
&\stackrel{(c)}{\leq} I(Y^n; W_1 | W_2) + 1 + P_e^{(n)} nR \\
&= H(Y^n | W_2) - H(Y^n | W_1, W_2) + 1 + P_e^{(n)} nR \\
&\stackrel{(d)}{=} \sum_{i=1}^n H(Y_i | Y^{i-1}, W_2) - \sum_{i=1}^n H(Y_i | W_1, W_2, Y^{i-1}) + 1 + P_e^{(n)} nR \\
&\stackrel{(e)}{=} \sum_{i=1}^n H(Y_i | Y^{i-1}, W_2, X_2^i) - \sum_{i=1}^n H(Y_i | W_1, W_2, Y^{i-1}, X_1^i, X_2^i) + 1 + P_e^{(n)} nR \\
&\stackrel{(f)}{\leq} \sum_{i=1}^n H(Y_i | Y^{i-1}, X_2^i) - \sum_{i=1}^n H(Y_i | Y^{i-1}, X_1^i, X_2^i) + 1 + P_e^{(n)} nR \\
&= \sum_{i=1}^n I(Y_i; X_1^i | Y^{i-1}, X_2^i) + 1 + P_e^{(n)} nR \\
&\leq I(X_1^n \rightarrow Y^n | X_2^n) + 1 + P_e^{(n)} nR,
\end{aligned} \tag{64}$$

where,

- (a) and (b) follow from the fact that the messages  $W_1$  and  $W_2$  are independent and chosen according to a uniform distribution,
- (c) follows from Fano's inequality,
- (d) follows from the chain rule,
- (e) follows from the fact that  $x_{1i}$  is a deterministic function given the message  $W_1$  and the feedback  $z_1^{i-1}$ , where the feedback  $z_1^{i-1}$  is a deterministic function of the output  $y^{i-1}$ ,
- (f) follows from the fact that the random variables  $W_1, W_2, X_1^i, X_2^i, Y^i$  form the Markov chain  $(W_1, W_2) - (X_1^i, X_2^i, Y^{i-1}) - Y_i$ .

Dividing (64) by  $n$ , we conclude that if there exists a code for which the error probability of decoding the messages  $W_1, W_2$  is  $P_e^{(n)}$  then the distribution  $Q(x_1^n | z_1^{n-1})Q(x_2^n | z_2^{n-1})$  induced by the code satisfies the first inequality of the outer bound theorem where  $\epsilon_n = \frac{1}{n} + P_e^{(n)} R$ . The proofs of the other two inequalities in (63) follow by a completely analogous sequence of steps as in (64): The proof of the second inequality of the outer bound starts with the equalities  $R_2 = H(W_2) = H(W_2 | W_1)$  and the third with  $R_1 + R_2 = H(W_1, W_2)$ . ■

*Corollary 20:* The outer bound given in Theorem 6 implies that  $\liminf \mathcal{R}_n$  is an outer bound for the achievable region.

*Proof:* Recall the definition of  $\mathcal{R}_n$  in eq. (20). Let  $(R_1, R_2)$  be an achievable rate pair. We will create a sequence of rate pairs  $(R_{1,n}, R_{2,n}) \in \mathcal{R}_n$  that converges to  $(R_1, R_2)$  and therefore, by the definition of  $\liminf$  of a sequence of sets (given in Appendix IV),  $(R_1, R_2) \in \liminf \mathcal{R}_n$ .

If  $(R_1, R_2) \in \mathcal{R}_n$  then we choose  $(R_{1,n}, R_{2,n}) = (R_1, R_2)$ . Otherwise we choose the closest point in  $\mathcal{R}_n$  to  $R_1, R_2$ . Because of inequality (63) the distance  $\|(R_{1,n}, R_{2,n}) - (R_1, R_2)\| \leq 2\epsilon_n$  and, therefore, the sequence  $(R_{1,n}, R_{2,n})$  converges to  $(R_1, R_2)$ . ■

### VIII. CAPACITY REGION OF THE FS-MAC WITHOUT FEEDBACK

The inner and outer bounds given in Theorems 5 and 6 specialize to the case where there is no feedback, i.e.,  $z_1, z_2$  are null. Hence, we can use it in order to extend Gallager's results [17, Ch. 4] on the capacity of indecomposable FSCs to indecomposable FS-MACs. An indecomposable FS-MAC (FSC) is a FS-MAC (FSC) for which the effect of the initial state vanishes with time. More precisely:

*Definition 3:* A FS-MAC (FSC) is *indecomposable* if, for every  $\epsilon > 0$ , there exists an  $n_0$  such that for  $n \geq n_0$ ,  $|P(s_n|x_1^n, x_2^n, s_0) - P(s_n|x_1^n, x_2^n, s'_0)| \leq \epsilon$  for all  $s_n, x_1^n, x_2^n, s_0$  and  $s'_0$ .

Since there is no feedback, according to Lemma 4 directed information becomes mutual information and causal conditioning becomes regular conditioning in all the expressions in the inner bound (Theorem 5) and outer bound (Theorem 6).

The proof of the capacity region of FS-MAC is based on the following two lemmas. The first lemma is used for showing that the difference between the lower bound and the upper bound goes to zero as  $n \rightarrow \infty$  and the second lemma, which is proved in Appendix V, is used for showing that the limits exist.

*Lemma 21:* Let  $\{Q(x_1^n)Q(x_2^n)\}_{n \geq 1}$  be an arbitrary sequence of input distribution. If the channel is an indecomposable FS-MAC then the following holds for all  $s'_0, s''_0$ :

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} |I(X_1^n; Y^n | X_2^n, s'_0) - I(X_1^n; Y^n | X_2^n, s''_0)| &= 0 \\ \lim_{n \rightarrow \infty} \frac{1}{n} |I(X_2^n; Y^n | X_1^n, s'_0) - I(X_2^n; Y^n | X_1^n, s''_0)| &= 0 \\ \lim_{n \rightarrow \infty} \frac{1}{n} |I(X_1^n, X_2^n; Y^n | s'_0) - I(X_1^n, X_2^n; Y^n | s''_0)| &= 0. \end{aligned} \quad (65)$$

*Proof:* The proof is identical to the proof of Theorem 4.6.4 in [17]. ■

The following lemma, which is proved in Appendix V, establishes the sup-additivity of  $\{\underline{\mathcal{R}}_n\}$ .

*Lemma 22:* (*sup-additivity of  $\underline{\mathcal{R}}_n$ .*) For any FS-MAC, the sequence  $\{\underline{\mathcal{R}}_n\}$ , which is defined in (19), is sup-additive, i.e.,

$$(n+l)\underline{\mathcal{R}}_{n+l} \supseteq n\underline{\mathcal{R}}_n + l\underline{\mathcal{R}}_l, \quad (66)$$

and therefore  $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n$  exists. Moreover, for an indecomposable FS-MAC without feedback  $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \mathcal{R}_n$  where  $\mathcal{R}_n$  is defined (20).

*Proof of Theorem 7:* Theorem 5 implies that  $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n$  is achievable, and Corollary 20 implies that  $\liminf_{n \rightarrow \infty} \mathcal{R}_n$  is an outer bound. Finally, since according to Lemma 22 the two limits are equal to  $\lim_{n \rightarrow \infty} \mathcal{R}_n$ , the capacity region is given by the last limit. ■

## IX. SUFFICIENT CONDITIONS FOR THE INNER AND OUTER BOUNDS TO COINCIDE FOR GENERAL FEEDBACK

## A. Stationary Finite state Markovian MAC with feedback

A stationary finite state Markovian MAC satisfies

$$P(y_i, s_i | x_{1i}, x_{2i}, s_{i-1}) = P(s_i | s_{i-1})P(y_i | s_{i-1}, x_{1i}, x_{2i}), \quad (67)$$

where the initial state distribution is the stationary distribution  $P(s_0)$ . In words, the states are not affected by the channel inputs.

For the stationary Markovian-MAC, the sequence  $\{\mathcal{R}_n\}$  is sup-additive. It follows from the fact that if we concatenate two input distributions  $Q_{n+k} = Q_n Q_k$ , then  $I(X_1^{n+k} \rightarrow Y^{n+k} | X_2^{n+k}) = I(X_1^n \rightarrow Y^n | X_2^n) + I(X_1^{n+k} \rightarrow Y^{n+k} | X_2^{n+k})$ , hence  $(n+k)\mathcal{R}_{n+k} \supseteq n\mathcal{R}_n + k\mathcal{R}_k$ . According to Lemma 23, the limit exists and is equal to

$$\lim_{n \rightarrow \infty} \mathcal{R}_n = \text{cl} \left( \bigcup_{n \geq 1} \mathcal{R}_n \right). \quad (68)$$

Next, we prove Theorem 8 that states that for a Markovian FS-MAC with a stationary ergodic state process, the inner bound (Theorem 5) and the outer bound (Theorem 6) coincide and therefore the capacity region is given by  $\lim_{n \rightarrow \infty} \mathcal{R}_n$ .

*Proof of Theorem 8:* Recall that the inner bound is given in Theorem 5 as  $\underline{\mathcal{R}}_N$  and the outer bound given in Theorem 6 and in Corollary 20 as  $\liminf \mathcal{R}_N$ . Next we show that the distance between  $\underline{\mathcal{R}}_N$  and  $\mathcal{R}_N$  goes to zero which implies by Lemma 25 that both limits equal and therefore the capacity region can be written as  $\lim \mathcal{R}_N$ .

Let us consider a specific input distribution denoted by  $\overline{Q}(x_1^N | z^{N-1})\overline{Q}(x_2^N | z^{N-1})$  corresponding to the region of the outer bound  $\mathcal{R}_N$ . Let us now consider an input distribution  $\underline{Q}$  for  $n+N$  inputs corresponding to the inner bound  $\underline{\mathcal{R}}_N$ , such that it is arbitrary for the first  $n$  inputs and then it is  $\overline{Q}(x_1^N | z^{N-1})\overline{Q}(x_2^N | z^{N-1})$ .

Now let us show that the term of the inner bound, i.e.  $I_{\underline{Q}}(X_1^N \rightarrow Y^N | X_2^{N+n}, s_0)$  and the term of the outer

bound  $I_{\underline{Q}}(X_1^N \rightarrow Y^N || X_2^N)$  are arbitrarily close to each other.

$$\begin{aligned}
& I_{\underline{Q}}(X_1^{N+n} \rightarrow Y^{N+n} || X_2^{N+n}, s_0) \\
& \stackrel{(a)}{\geq} I_{\underline{Q}}(X_1^{N+n} \rightarrow Y^{N+n} || X_2^{N+n}, S_n, s_0) - \log |\mathcal{S}| \\
& \stackrel{(b)}{\geq} \sum_{i=n+1}^{N+n} H_{\underline{Q}}(Y_i | Y^{i-1}, X_2^i, S_n, s_0) - H_{\underline{Q}}(Y_i | Y^{i-1}, X_2^i, X_1^i, S_n, s_0) - \log |\mathcal{S}| \\
& \stackrel{(c)}{\geq} \sum_{i=n+1}^{N+n} H_{\underline{Q}}(Y_i | Y_{n+1}^{i-1}, X_{2,n+1}^i, S_n, s_0) - H_{\underline{Q}}(Y_i | Y_{n+1}^{i-1}, X_{2,n+1}^i, X_{1,n+1}^i, S_n, s_0) - \log |\mathcal{S}| \\
& = I_{\underline{Q}}(X_{1,n+1}^{N+n} \rightarrow Y_{n+1}^{N+n} || X_{2,n+1}^{N+n}, S_n, s_0) - H(S_n) \\
& \stackrel{(d)}{\geq} I_{\underline{Q}}(X_{1,n+1}^{N+n} \rightarrow Y_{n+1}^{N+n} || X_{2,n+1}^{N+n}, S_n)(1 - \delta) - \log |\mathcal{S}| \\
& \geq I_{\underline{Q}}(X_{1,n+1}^{N+n} \rightarrow Y_{n+1}^{N+n} || X_{2,n+1}^{N+n}, S_n) - \delta(N + n) \log |\mathcal{Y}| - \log |\mathcal{S}| \\
& \stackrel{(e)}{\geq} I_{\underline{Q}}(X_{1,n+1}^{N+n} \rightarrow Y_{n+1}^{N+n} || X_{2,n+1}^{N+n}) - \delta(N + n) \log |\mathcal{Y}| - 2 \log |\mathcal{S}| \\
& \stackrel{(f)}{\geq} I_{\underline{Q}}(X_1^N \rightarrow Y^N || X_2^N) - \delta(N + n) \log |\mathcal{Y}| - 2 \log |\mathcal{S}|, \tag{69}
\end{aligned}$$

where

- (a) follows from Lemma 2 that states that conditioning on  $S_n$  can differ at most by  $\log |\mathcal{S}|$ ,
- (b) follows from omitting the first  $n$  elements in the sum that defines directed information,
- (c) follows from the fact that conditioning decreases entropy,
- (d) follows from the fact that the Markov chain is ergodic, hence for any  $\delta > 0$ , there exists an  $n$  such that  $|P(s_n | s_0) - P(s_n)| \leq \delta$  for any  $s_0 \in \mathcal{S}$  and  $s_n \in \mathcal{S}$ , where  $P(s_n)$  is the stationary distribution of  $s_n$ ,
- (e) follows from Lemma 2 that states that conditioning on  $S_n$  can differ by at most  $\log |\mathcal{S}|$ ,
- (f) follows from the stationarity of the channel.

Dividing both sides by  $N + n$  we get that for any  $s_0$ ,

$$\frac{1}{N + n} I_{\underline{Q}}(X_1^{N+n} \rightarrow Y^{N+n} || X_2^{N+n}, s_0) - \frac{1}{N + n} I_{\underline{Q}}(X_1^N \rightarrow Y^N || X_2^N) \geq -\delta(1 + \frac{n}{N}) \log |\mathcal{Y}| - 2 \frac{\log |\mathcal{S}|}{N + n} \tag{70}$$

Inequality (70) shows that the difference between the upper bound region and the lower bound is arbitrarily small for  $N$  large enough and, hence, in the limit the regions coincide.  $\blacksquare$

### B. Finite State Markovian MAC with limited ISI

In this subsection we consider a MAC inspired by Kim's point-to-point channel [22]. The conditional probability of the MAC is given by

$$P(y_i, z_i | x_1^i, x_2^i, z_{i-1}) = P(z_i | z_{i-1}) P(y_i | z_{i-1}, x_{1,i-m}^i, x_{2,i-m}^i), \quad i = 1, 2, 3, \dots \tag{71}$$

where the distribution of  $Z_0$  is the stationary distribution  $P(z_0)$ , and there is also some initial distribution  $P(x_{-m+1}, \dots, x_0)$ .

This channel is a FS-MAC where the state at time  $i$  is  $(z_{i-1}, x_{1,i-m}^{i-1}, x_{2,i-m}^{i-1})$  and therefore the inner bound (Theorem 5) and the outer bound (Theorem 6) apply to this channel. Theorem 8 also holds for this kind of channels, namely, the capacity region is given by  $\lim_{n \rightarrow \infty} \mathcal{R}_n$ . The proof is very similar, the only difference being that the input  $\underline{Q}$  for  $n + N$  inputs is constructed slightly differently: it is arbitrary for the first  $n - m$  inputs, then it is as the initial distribution  $P(x_{-m+1}, \dots, x_0)$ , and then it is  $\overline{Q}(x_1^N || z^{N-1}) \overline{Q}(x_2^N || z^{N-1})$ .

It is also possible to represent the channel with an alternative law, identical to the law of the channel given in eq. (71) for  $i \geq m + 1$  but for  $i \leq m$  the output  $y_i$  is not influenced by the input and is, with probability 1, a particular output  $\phi \in \mathcal{Y}$ . Let us define  $\mathcal{R}_n^\phi$  similarly as  $\mathcal{R}_n$  but with the alternative law for the channel. On one hand, it is clear that  $\mathcal{R}_n^\phi \subseteq \mathcal{R}_n$  for all  $n$ , and on the other hand the difference between  $\mathcal{R}_n^\phi$  and  $\mathcal{R}_n$  is at most  $m \log \mathcal{Y}$  because it is possible to use the distribution of the first  $m$  inputs,  $Q(x_1^m)$ , to create a desired initial distribution and then use the same input as in  $\mathcal{R}_n$ . Hence,

$$\lim_{n \rightarrow \infty} \mathcal{R}_n^\phi = \lim_{n \rightarrow \infty} \mathcal{R}_n. \quad (72)$$

The advantage of analyzing  $\mathcal{R}_n^\phi$  rather than analyzing  $\mathcal{R}_n$  is that the sequence  $n\mathcal{R}_n^\phi$  is sup-additive, i.e.  $(n+l)\mathcal{R}_{n+l}^\phi \supseteq n\mathcal{R}_n^\phi + l\mathcal{R}_l^\phi$ , and according to Lemma 23,  $\lim_{n \rightarrow \infty} \mathcal{R}_n^\phi = \text{cl} \left( \bigcup_{n \geq 1} \mathcal{R}_n^\phi \right)$ . Hence, we can conclude that Theorem 9 holds for this channel too, namely, if the capacity of the Finite state Markovian MAC with limited ISI is zero without feedback then it is zero also in the presence of feedback.

## X. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper we have shown that directed information and causal conditioning emerge naturally in characterizing the capacity region of FS-MACs in the presence of a time-invariant feedback. The capacity region is given as a ‘multi-letter’ expression and it is a first step toward deriving useful concepts in communication. For instance, we use this characterization in order to show that for a stationary and ergodic Markovian channel, the capacity is zero if and only if the capacity with feedback is zero. Further, we identify FS-MACs for which feedback does not enlarge the capacity region and for which source-channel separation holds.

For the point-to-point channel with feedback, recent work has shown that, for some families of channels such as unifilar channels [28] or the additive Gaussian where the noise is ARMA [22], the directed information formula can be computed and, further, can lead to the development of capacity achieving coding schemes. One future direction is to use the characterizations developed in this paper to explicitly compute the capacity regions of classes of MACs with memory and feedback (other than the multiplexer followed by a point-to-point channel), and to find optimal coding schemes.



APPENDIX I  
PROOF OF LEMMA 3

Recall that Lemma 3 states that if

$$Q(x_1^N, x_2^N || y^{N-1}) = Q(x_1^N || y^{N-1})Q(x_2^N || y^{N-1}), \quad (73)$$

then

$$\mathcal{I}(Q(x_1^N, x_2^N || y^{N-1}); P(y^N || x_1^N, x_2^N)) = I(X_1^N \rightarrow Y^N || X_2^N). \quad (74)$$

*Proof:* The following sequence of equalities proves the lemma.

$$\begin{aligned}
& \mathcal{I}(Q(x_1^N, x_2^N || y^{N-1}); P(y^N || x_1^N, x_2^N)) \\
& \stackrel{(a)}{=} \mathcal{I}(Q(x_1^N || y^{N-1})Q(x_2^N || y^{N-1}); P(y^N || x_1^N, x_2^N)) \\
& \stackrel{(b)}{=} \sum_{y^N, x_1^N, x_2^N} Q(x_1^N || y^{N-1})Q(x_2^N || y^{N-1})P(y^N || x_1^N, x_2^N) \frac{P(y^N || x_1^N, x_2^N)}{\sum_{x_1'^N} Q(x_1'^N || y^{N-1})P(y^N || x_1'^N, x_2^N)} \\
& \stackrel{(c)}{=} \sum_{y^N, x_1^N, x_2^N} P(x_1^N, x_2^N, y^N) \frac{P(y^N || x_1^N, x_2^N)}{\sum_{x_1'^N} Q(x_1'^N || y^{N-1})P(y^N || x_1'^N, x_2^N)} \\
& = \mathbf{E} \left[ \frac{P(y^N || x_1^N, x_2^N)}{\sum_{x_1'^N} Q(x_1'^N || y^{N-1}, x_2^N)P(y^N || x_1'^N, x_2^N)} \right] \\
& = \mathbf{E} \left[ \frac{Q(x_2^N || y^{N-1})P(y^N || x_1^N, x_2^N)}{Q(x_2^N || y^{N-1}) \sum_{x_1'^N} Q(x_1'^N || y^{N-1}, x_2^N)P(y^N || x_1'^N, x_2^N)} \right] \\
& = \mathbf{E} \left[ \frac{Q(x_2^N || y^{N-1})P(y^N || x_1^N, x_2^N)}{\sum_{x_1'^N} P(y^N, x_1'^N, x_2^N)} \right] \\
& = \mathbf{E} \left[ \frac{Q(x_2^N || y^{N-1})P(y^N || x_1^N, x_2^N)}{P(x_2^N, y^N)} \right] \\
& = \mathbf{E} \left[ \frac{P(y^N || x_1^N, x_2^N)}{P(y^N || x_2^N)} \right] \\
& \stackrel{(d)}{=} I(X_1^N \rightarrow Y^N || X_2^N)
\end{aligned} \quad (75)$$

- (a) follows from the assumption given in eq. (73).
- (b) follows from the definition of the functional  $\mathcal{I}(Q; P)$  given in eq. (16).
- (c) follows from Lemma 1 that states that  $P(x_1^N, x_2^N, y^N) = Q(x_1^N, x_2^N || y^{N-1})P(y^N || x_1^N, x_2^N)$  and the assumption given in (73).
- (d) follows from the definition of directed information.

■

## APPENDIX II

### PROOF OF LEMMA 4

Lemma 4 states that if

$$Q(x_1^N, x_2^N || y^{N-1}) = Q(x_1^N)Q(x_2^N), \quad (76)$$

then

$$I(X_1^N; Y^N | X_2^N) = I(X_1^N \rightarrow Y^N | X_2^N). \quad (77)$$

*Proof:* The following sequence of equalities proves the lemma.

$$\begin{aligned} I(X_1^N; Y^N | X_2^N) &= \mathbf{E} \left[ \log \frac{P(Y^N, X_1^N | X_2^N)}{P(Y^N | X_2^N) Q(X_1^N | X_2^N)} \right] \\ &\stackrel{(a)}{=} \mathbf{E} \left[ \log \frac{P(Y^N, X_1^N, X_2^N)}{P(Y^N, X_2^N) Q(X_1^N | X_2^N)} \right] \\ &\stackrel{(b)}{=} \mathbf{E} \left[ \log \frac{Q(X_1^N, X_2^N || Y^{N-1}) P(Y^N || X_1^N, X_2^N)}{P(Y^N || X_2^N) Q(X_2^N || Y^{N-1}) Q(X_1^N | X_2^N)} \right] \\ &\stackrel{(c)}{=} \mathbf{E} \left[ \log \frac{Q(X_1^N) Q(X_2^N) P(Y^N || X_1^N, X_2^N)}{P(Y^N || X_2^N) Q(X_2^N) Q(X_1^N)} \right] \\ &= \mathbf{E} \left[ \log \frac{P(Y^N || X_1^N, X_2^N)}{P(Y^N || X_2^N)} \right] \\ &= I(X_1^N \rightarrow Y^N | X_2^N). \end{aligned} \quad (78)$$

- (a) follows from multiplying the numerator and denominator by  $P(x_2^N)$ .
- (b) follows from decomposing the joint distributions  $P(y^N, x_1^N, x_2^N)$  and  $P(Y^N, X_2^N)$  into causal conditioning distribution by using Lemma 1.
- (c) follows from the fact that the assumption of the lemma given in (76) implies that  $Q(X_1^N, X_2^N) = Q(X_1^N)Q(X_2^N)$ . This can be obtained by multiplying both sides of (76) by  $P(y^n || x_1^n, x_2^n)$  and then summing over all  $y^n \in \mathcal{Y}^n$ .

■

## APPENDIX III

### PROOF OF LEMMA 12

Lemma 12 states that

$$\max_{Q(x_1^n || y^{n-1}) Q(x_2^n || y^{n-1})} I(X_1^n, X_2^n \rightarrow Y^n) = 0 \iff \max_{Q(x_1^n) Q(x_2^n)} I(X_1^n, X_2^n \rightarrow Y^n) = 0, \quad (79)$$

and each condition also implies that  $P(y^n || x_1^n, x_2^n) = P(y^n)$  for all  $x_1^n, x_2^n$ .

*Proof:* Proving the direction  $\implies$  is trivial since

$$\max_{Q(x_1^n || y^{n-1}) Q(x_2^n || y^{n-1})} I(X_1^n, X_2^n \rightarrow Y^n) \geq \max_{Q(x_1^n) Q(x_2^n)} I(X_1^n, X_2^n \rightarrow Y^n). \quad (80)$$

For the other direction,  $\Leftarrow$ , we have the assumption that  $I(X_1^n, X_2^n \rightarrow Y^n) = 0$  for all input distributions  $Q(x_1^n)Q(x_2^n)$ , and in particular for the case that  $X_1^n$  and  $X_2^n$  are uniformly distributed over their alphabets. Directed information can be written as a Kullback Leibler divergence, i.e.,

$$\sum_{x_1^n, x_2^n, y^n} Q(x_1^n)Q(x_2^n)P(y^n|x_1^n, x_2^n) \log \frac{Q(x_1^n)Q(x_2^n)P(y^n|x_1^n, x_2^n)}{P(y^n)Q(x_1^n)Q(x_2^n)} = 0 \quad (81)$$

and by using the fact that if the Kullback Leibler divergence  $D(P||Q) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}$  is zero, then  $P(x) = Q(x)$  for all  $x \in \mathcal{X}$ , we conclude that (81) implies that  $P(y^n|x_1^n, x_2^n) = P(y^n)$  for all  $x_1^n \in \mathcal{X}_1^n$  and all  $x_2^n \in \mathcal{X}_2^n$ . It follows that

$$\begin{aligned} \max_{Q(x_1^n|y^{n-1})Q(x_2^n|y^{n-1})} I(X_1^n, X_2^n \rightarrow Y^n) &= \max_{Q(x_1^n|y^{n-1})Q(x_2^n|y^{n-1})} \mathbf{E} \left[ \log \frac{P(Y^n|X_1^n, X_2^n)}{P(Y^n)} \right] \\ &= \max_{Q(x_1^n|y^{n-1})Q(x_2^n|y^{n-1})} \mathbf{E}[0] = 0. \end{aligned} \quad (82)$$

■

## APPENDIX IV

### SUP-ADDITIVITY AND CONVERGENCE OF 2D REGIONS

Let  $A, B$  be sets in  $\mathbb{R}^2$ , i.e.,  $A$  and  $B$  are sets of 2D vectors. The sum of two regions is denoted as  $A + B$  and defined as

$$A + B = \{\mathbf{a} + \mathbf{b} : \mathbf{a} \in A, \mathbf{b} \in B\}, \quad (83)$$

and multiplication of a set  $A$  with a scalar  $c$  is defined as

$$cA = \{c\mathbf{a} : \mathbf{a} \in A\}. \quad (84)$$

A sequence  $\{A_n\}$ ,  $n = 1, 2, 3, \dots$ , of 2D regions is said to *converge* to a region  $A$ , written  $A = \lim A_n$  if

$$\limsup A_n = \liminf A_n = A \quad (85)$$

where

$$\begin{aligned} \liminf A_n &= \{\mathbf{a} : \mathbf{a} = \lim \mathbf{a}_n, \mathbf{a}_n \in A_n\}, \\ \limsup A_n &= \{\mathbf{a} : \mathbf{a} = \lim \mathbf{a}_k, \mathbf{a}_k \in A_{n_k}\}, \end{aligned} \quad (86)$$

and  $n_k$  denotes an arbitrary increasing subsequence of the integers. An alternative and equivalent definition of  $\limsup$  and  $\liminf$  is given by  $\limsup A_n = \bigcap_{n \geq 1} \text{cl} \left( \bigcup_{m \geq n} A_m \right)$  and  $\liminf A_n = \bigcup_{n \geq 1} \text{cl} \left( \bigcap_{m \geq n} A_m \right)$ . For more details on convergence of sets in finite dimensions see [48].

Let  $\overline{A}$  denote

$$\overline{A} = \text{cl} \left( \bigcup_{n \geq 1} A_n \right). \quad (87)$$

We say that a sequence  $\{A_n\}_{n \geq 1}$  is *bounded* if  $\sup\{\|\mathbf{a}\| : \mathbf{a} \in \overline{A}\} < \infty$  where  $\|\cdot\|$  denotes a norm in  $\mathbb{R}^2$ .

*Lemma 23:* Let  $A_n$ ,  $n = 1, 2, \dots$ , be a bounded sequence of sets in  $\mathbb{R}^2$  that includes the origin, i.e.  $(0, 0)$ . If  $nA_n$  is sup-additive, i.e., for all  $n \geq 1$  and all  $N > n$

$$NA_N \supseteq nA_n + (N - n)A_{N-n} \quad (88)$$

then

$$\lim_{n \rightarrow \infty} A_n = \overline{A}. \quad (89)$$

*Proof:* From the definitions we have  $\overline{A} \supseteq \limsup A_n \supseteq \liminf A_n$ . Hence it is enough to show that  $\overline{A} \subseteq \liminf A_n$ .

Let  $\mathbf{a}$  be a point in  $\overline{A}$ . Then for every  $\epsilon > 0$  there exists an  $n$  and a point  $\mathbf{a}_\epsilon$  such that  $\mathbf{a}_\epsilon \in A_n$  and  $\|\mathbf{a} - \mathbf{a}_\epsilon\| \leq \epsilon$ . By induction we prove that for any integer  $m \geq 2$ ,  $A_n \subseteq A_{mn}$ , and this implies that  $\mathbf{a}_\epsilon \in A_{mn}$ . For  $m = 2$  we choose  $N = 2n$  and we get that

$$A_{2n} \supseteq \frac{A_n}{2} + \frac{A_n}{2} \supseteq A_n. \quad (90)$$

Now assume that it holds for  $m - 1$  and let us show that it holds for  $m$ .

$$A_{mn} \supseteq \frac{A_n}{m} + \frac{(m-1)A_{(m-1)n}}{m} \supseteq \frac{A_n}{m} + \frac{(m-1)A_n}{m} \supseteq A_n. \quad (91)$$

Now, for any  $N > n$ , we can represent  $N$  as  $mn + j$  where  $0 \leq j \leq n - 1$ , hence

$$A_{mn+j} \supseteq \frac{j}{mn+j}A_j + \frac{mn}{mn+j}A_{mn}. \quad (92)$$

Because  $\mathbf{a}_\epsilon$  is in  $A_n$ , then it implies that it is in  $A_{mn}$  too. Following (92) and the fact that  $(0, 0) \in A_j$  we obtain

$$\frac{mn}{mn+j}\mathbf{a}_\epsilon \in A_{mn+j}. \quad (93)$$

For any  $\delta > 0$  and for any  $N \geq \frac{n}{\delta}$  we conclude the existence of an element in  $A_N$  for which the distance from  $\mathbf{a}$  can be upper-bounded by

$$\left\| \frac{mn}{mn+j}\mathbf{a}_\epsilon - \mathbf{a} \right\| = \left\| \mathbf{a}_\epsilon - \mathbf{a} - \frac{j}{mn+j}\mathbf{a}_\epsilon \right\| \leq \|\mathbf{a}_\epsilon - \mathbf{a}\| + \delta \|\mathbf{a}_\epsilon\| \leq \epsilon + \delta \|\mathbf{a}_\epsilon\|. \quad (94)$$

Because  $\epsilon$  and  $\delta$  are arbitrarily small we can find a sequence of points  $\mathbf{a}_n \in A_n$  that converges to  $\mathbf{a}$  and therefore  $\mathbf{a} \in \liminf A_n$ , which implies that  $\overline{A} \subseteq \liminf A_n$ .  $\blacksquare$

*Corollary 24:* For a sup-additive sequence, as defined in Lemma 23, the limit is convex.

This corollary follows immediately from the definition of the sup-additivity property, eq. (88) where  $n = \alpha N$ , where  $0 < \alpha < 1$ , and  $N$  goes to infinity.

The (*Hausdroff*) distance between two sets  $A$  and  $B$ , is defined as

$$d(A, B) = \max\{\sup[d(\mathbf{a}, B) : \mathbf{a} \in A], \sup[d(\mathbf{b}, A) : \mathbf{b} \in B]\}, \quad (95)$$

where the distance between a set  $A$  and a point  $\mathbf{b}$  is given by,

$$d(\mathbf{b}, A) = \inf_{\mathbf{a}} \|\mathbf{a} - \mathbf{b}\| : \mathbf{a} \in A \quad (96)$$

*Lemma 25:* If  $\lim_{n \rightarrow \infty} d(A_n, B_n) = 0$  then

$$\begin{aligned} \limsup A_n &= \limsup B_n, \\ \liminf A_n &= \liminf B_n. \end{aligned} \tag{97}$$

*Proof:* The proof is straightforward. Given a sequence  $\{\mathbf{a}_k\} \in A_{n_k}$  that converges to  $\mathbf{a}$ , we construct a sequence  $\{\mathbf{b}_k\}$  by finding a point in  $B_{n_k}$  that is at a distance less than  $\frac{1}{k} + d(\mathbf{a}_k, B_{n_k})$ . Since the distance between the sets goes to zero,  $\lim \mathbf{b}_k = \lim \mathbf{a}_k = \mathbf{a}$  and from the definitions of limits of sets, it implies that (97) holds. ■

## APPENDIX V

### PROOF OF LEMMA 22

Recall the definition of  $\underline{\mathcal{R}}_n$  and  $\mathcal{R}_n$  in (19) and (20) respectively.

Lemma 22 states that

$$(n+l)\underline{\mathcal{R}}_{n+l} \supseteq n\underline{\mathcal{R}}_n + l\underline{\mathcal{R}}_l. \tag{98}$$

and for an indecomposable FS-MAC without feedback  $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \mathcal{R}_n$ .

*Proof of Lemma 22:* We notice that if a sequence of sets is sup-additive then the sequence of the convex hull of the sets is also sup-additive. Hence, it is enough to prove the sup-additivity of the sequence  $\underline{\mathcal{R}}_n$  without the appearance of the random variable  $W$  that its role is to convexify the regions.

The set  $\underline{\mathcal{R}}_n$  is defined by three expressions that involve directed information. Because each expression is sup-additive the whole set is sup-additive. We prove that the first expression, i.e.  $\min_{s_0} I(X_1^n \rightarrow Y^n | X_2^n, s_0) - \log |\mathcal{S}|$  is sup-additive (the proofs of the sup-additivity of the other expressions are similar and therefore omitted).

$$\begin{aligned} & \min_{s_0} I(X_1^{n+l} \rightarrow Y^{n+l} | X_2^{n+l}, s_0) \\ & \stackrel{(a)}{\geq} \min_{s_0} \sum_{i=1}^n I(Y_i; X_1^i | Y^{i-1}, X_2^i, s_0) + \min_{s_0} \sum_{j=n+1}^{n+l} I(Y_j; X_1^j | Y^{j-1}, X_2^j, s_0) \\ & \stackrel{(b)}{\geq} I(X_1^n \rightarrow Y^n | X_2^n, s_0) + \sum_{j=n+1}^{n+l} I(Y_j; X_{1,n+1}^j | Y^{j-1}, X_{2,n+1}^j, s_0) \\ & \stackrel{(c)}{\geq} I(X_1^n \rightarrow Y^n | X_2^n, s_0) + \sum_{j=n+1}^{n+l} I(Y_j; X_{1,n+1}^j | Y^{j-1}, X_{2,n+1}^j, s_n) - \log |\mathcal{S}| \\ & = \min_{s_0} I(X_1^n \rightarrow Y^n | X_2^n, s_0) + \min_{s_0} \sum_{s_n} P(s_n | s_0) \sum_{j=n+1}^{n+l} I(Y_j; X_{1,n+1}^j | Y^{j-1}, X_{2,n+1}^j, s_n) - \log |\mathcal{S}| \\ & \geq \min_{s_0} I(X_1^n \rightarrow Y^n | X_2^n, s_0) + \min_{s_n} \sum_{j=n+1}^{n+l} I(Y_j; X_{1,n+1}^j | Y_{n+1}^{j-1}, X_{2,n+1}^j, s_n) - \log |\mathcal{S}| \\ & \stackrel{(d)}{=} \min_{s_0} I(X_1^n \rightarrow Y^n | X_2^n, s_0) + \min_{s_0} I(X_1^l \rightarrow Y^l | X_2^l, s_0) - \log |\mathcal{S}|. \end{aligned} \tag{99}$$

(a) follows the definition of the directed information the fact that  $\min_s [f(s) + g(s)] \geq \min_s f(s) + \min_s g(s)$ ,

- (b) follows the fact that  $I(X; Y, Z) \geq I(X; Y)$ ,
- (c) follows Lemma 2 that states that conditioning by  $S_n$  can differ by at most  $\log |S|$ ,
- (d) follows from the stationarity of the channel.

According to Lemma 23, since the sequence  $\{\underline{\mathcal{R}}_n\}$  is sup-additive the limit exists. In the rest of the proof we show that  $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \mathcal{R}_n$ . The terms of the region  $\underline{\mathcal{R}}_n$  have an auxiliary random variable  $W$  whose only role is to convexify the region. Let us denote  $\underline{\mathcal{R}}_n^o$  the same region as  $\underline{\mathcal{R}}_n$  where  $W$  is restricted to be null. We show first that restricting  $W$  to being null does not influence the limit, i.e.,  $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n^o$ . In the first half of the proof we showed that  $\underline{\mathcal{R}}_n^o$  is sub-additive. Using this fact, we show now, that any convex combination with rational weights  $(\frac{l}{k}, \frac{k-l}{k})$  of any two points from  $\underline{\mathcal{R}}_n^o$  is in  $\underline{\mathcal{R}}_{kn}^o$ .

$$\underline{\mathcal{R}}_{kn}^o \supseteq \frac{l}{k} \underline{\mathcal{R}}_{ln}^o + \frac{k-l}{k} \underline{\mathcal{R}}_{(k-l)n}^o \supseteq \frac{l}{k} \underline{\mathcal{R}}_n^o + \frac{k-l}{k} \underline{\mathcal{R}}_n^o \quad (100)$$

The left and the right inclusions in (100) are due to the sup-additivity of  $\underline{\mathcal{R}}_n^o$ . The left inclusion is from the definition of the sup-additivity and the right is due to the fact that sup-additivity of  $\underline{\mathcal{R}}_n^o$  also implies that for any two positive integers  $m, n$ ,  $\underline{\mathcal{R}}_{mn}^o \supseteq \underline{\mathcal{R}}_n^o$  (This is shown by induction in (90,91)). From (100) we can deduce that for any  $\epsilon > 0$  we can find a  $k(\epsilon)$  such that  $\underline{\mathcal{R}}_n \subseteq \underline{\mathcal{R}}_{nk}^o + \epsilon$ . This fact, together with the trivial fact that  $\underline{\mathcal{R}}_n \supseteq \underline{\mathcal{R}}_n^o$ , and the fact that the limits of both sequences exist, allow us to deduce that the limits are the same, i.e.,  $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n^o$ .

We conclude the proof by showing that, for any input distribution  $Q(x_1^n)Q(x_2^n)$ , the difference between the terms in the inequalities of  $\{\underline{\mathcal{R}}_n^o\}$  and  $\{\mathcal{R}_n\}$  goes to zero as  $n \rightarrow \infty$ , hence the distance between the sets of the sequences goes to zero as  $n \rightarrow \infty$  and, by Lemma 25, the limits of the sequences are the same.

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \left| I(X_1^n \rightarrow Y^n | X_2^n) - \min_{s_0} I(X_1^n \rightarrow Y^n | X_2^n, s_0) + \log |S| \right| \\ & \stackrel{(a)}{\leq} \lim_{n \rightarrow \infty} \frac{1}{n} \left| I(X_1^n \rightarrow Y^n | X_2^n, S_0) - \min_{s_0} I(X_1^n \rightarrow Y^n | X_2^n, s_0) + \log |S| \right| + \log |S| \\ & = \lim_{n \rightarrow \infty} \frac{1}{n} \left[ I(X_1^n \rightarrow Y^n | X_2^n, S_0) - \min_{s_0} I(X_1^n \rightarrow Y^n | X_2^n, s_0) \right] \\ & \stackrel{(b)}{\leq} \lim_{n \rightarrow \infty} \frac{1}{n} \left[ \max_{s_0} I(X_1^n \rightarrow Y^n | X_2^n, s_0) - \min_{s_0} I(X_1^n \rightarrow Y^n | X_2^n, s_0) \right] \\ & \stackrel{(c)}{=} 0 \end{aligned} \quad (101)$$

- (a) follows from Lemma 2 and the triangle inequality.
- (b) follows from the fact that  $\max_{s_0} I(X_1^n \rightarrow Y^n | X_2^n, s_0) \geq I(X_1^n \rightarrow Y^n | X_2^n, S_0)$ .
- (c) follows from Lemma 21 that states this equality for indecomposable FS-MAC without feedback (recall also that directed information equals mutual information in the absence of feedback).

■

APPENDIX VI  
PROOF OF THEOREM 16

$$\begin{aligned}\mathbf{E}[P_{e1}] &= \sum_{y^N} \sum_{x_1^N, x_2^N} P(x_1^N, x_2^N, y^N) P[\text{error1} | m_1, m_2, x_1^N, x_2^N, y^N] \\ &= \sum_{y^N} \sum_{x_1^N, x_2^N} Q(x_1^N | z_1^{N-1}) Q(x_2^N | z_2^{N-1}) P(y^N | x_1^N, x_2^N) P[\text{error1} | m_1, m_2, x_1^N, x_2^N, y^N],\end{aligned}\quad (102)$$

where  $P[\text{error1} | m_1, m_2, x_1^N, x_2^N, y^N]$  is the error probability of decoding  $m_1$  given that  $m_2$  is decoded correctly. Throughout the remainder of the proof we fix the message  $m_1, m_2$ . For a given tuple  $(m_1, m_2, x_1^N, x_2^N, y^N)$  define the event  $A_{m'_1}$ , for each  $m'_1 \neq m_1$ , as the event that the message  $m'_1$  is selected in such a way that  $P(y^N | m'_1, m_2) > P(y^N | m, m_2)$  which is the same as  $P(y^N | x'_1{}^N, x_2^N) > P(y^N | x_1^N, x_2^N)$  where  $x'_1{}^N$  is a shorthand notation for  $x_1^N(m'_1, z_1^{N-1}(y^{N-1}))$  and  $x_i^N$  is a shorthand notation for  $x_i^N(m_l, z_l^{N-1}(y^{N-1}))$  for  $l = 1, 2$ . From the definition of  $A_{m'_1}$  we have

$$\begin{aligned}P(A_{m'_1} | m_1, m_2, x_1^N, x_2^N, y^N) &= \sum_{x'^N} Q(x'_1{}^N | z^{N-1}) \cdot \mathbf{1}[P(y^N | x'_1{}^N, x_2^N) > P(y^N | x_1^N, x_2^N)] \\ &\leq \sum_{x'^N} Q(x'_1{}^N | z^{N-1}) \left[ \frac{P(y^N | x'_1{}^N, x_2^N)}{P(y^N | x_1^N, x_2^N)} \right]^s; \quad \text{any } s > 0\end{aligned}\quad (103)$$

where  $\mathbf{1}(x)$  denotes the indicator function.

$$\begin{aligned}P[\text{error1} | m_1, m_2, x_1^N, x_2^N, y^N] &= P\left(\bigcup_{m'_1 \neq m} A_{m'_1} | m_1, m_2, x_1^N, x_2^N, y^N\right) \\ &\leq \min \left\{ \sum_{m'_1 \neq m} P(A_{m'_1} | m_1, m_2, x_1^N, x_2^N, y^N), 1 \right\} \\ &\leq \left[ \sum_{m'_1 \neq m_1} P(A_{m'_1} | m_1, m_2, x_1^N, x_2^N, y^N) \right]^\rho; \quad \text{any } 0 \leq \rho \leq 1 \\ &\leq \left[ (M_1 - 1) \sum_{x'_1{}^N} Q(x'_1{}^N | z^{N-1}) \left[ \frac{P(y^N | x'_1{}^N, x_2^N)}{P(y^N | x_1^N, x_2^N)} \right]^s \right]^\rho, \quad 0 \leq \rho \leq 1, s > 0,\end{aligned}\quad (104)$$

where the last inequality is due to inequality (103). By substituting inequality (104) in eq. (102) we obtain:

$$\mathbf{E}[P_{e1}] \leq (M - 1)^\rho \sum_{y^N, x_2^N} Q(x_2^N | z_2^{N-1}) \left[ \sum_{x^N} Q(x_1^N | z_1^{N-1}) P(y^N | x_1^N, x_2^N)^{1-s\rho} \right] \left[ \sum_{x'_1{}^N} Q(x'_1{}^N | z^{N-1}) P(y^N | x'_1{}^N, x_2^N)^s \right]^\rho$$

By substituting  $s = 1/(1 + \rho)$ , and recognizing that  $x'$  is a dummy variable of summation, we obtain eq. (37) and complete the proof of the bound on  $\mathbf{E}[P_{e1}]$ .

The proof for bounding  $\mathbf{E}[P_{e2}]$  is identical to the proof that is given here for  $\mathbf{E}[P_{e1}]$ , up to exchanging the indices. For  $\mathbf{E}[P_{e3}]$  the upper bound is identical to the case of the point-to-point channel with an input  $x_1^N, x_2^N$ , as proven in [27] where the union bound which appears here in eq. (104) consists of  $(M_1 - 1)(M_2 - 1)$  terms. ■

## REFERENCES

- [1] R. Ahlswede, "Multi-way communication channels," in *Proceedings of 2nd International Symposium on Information Theory (Thakadsor, Armenian SSR, Sept. 1971)*, 1973, pp. 23–52.
- [2] T. M. Cover and C. S. K. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. on Info. Theory*, vol. 27, no. 3, pp. 292–298, 1981.
- [3] F. M. J. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels," *IEEE Trans. on Info. Theory*, vol. 28, no. 1, pp. 93–95, 1982.
- [4] S. Bross and A. Lapidoth, "An improved achievable region for the discrete memoryless two-user multiple-access channel with noiseless feedback," *IEEE Trans. on Info. Theory*, vol. 51, pp. 811–833, 2005.
- [5] W. Wu, S. Vishwanath, and A. Arapostathis, "On the capacity of multiple access channels with state information and feedback," 2006, submitted to *IEEE Trans. Inform. Theory*. Available at [arxiv.org/cs.IT/0606014](http://arxiv.org/cs.IT/0606014).
- [6] L. H. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 623–628, 1984.
- [7] J. P. M. Schalkwijk and T. Kailath, "Coding scheme for additive noise channels with feedback I: No bandwidth constraint," *IEEE Trans. Inform. Theory*, vol. 12, pp. 172–182, 1966.
- [8] A. Lapidoth and M. A. Wigger, "On the Gaussian mac with imperfect feedback," in *24th IEEE Convention of Electrical and Electronics Engineers in Israel (IEEEI06), Eilat, Israel*, November 2006.
- [9] N. Martins and T. Weissman, "Coding schemes for additive white noise channels with feedback corrupted by quantization or bounded noise," 2006, submitted to *IEEE Trans. Inform. Theory*. Available at [arxiv.org/cs.IT/0609055](http://arxiv.org/cs.IT/0609055).
- [10] N. Merhav and T. Weissman, "Coding for the feedback Gel'fand-Pinsker channel and the feedforward Wyner-Ziv source," *IEEE Trans. on Info. Theory*, vol. 52, pp. 4207–4211, 2006.
- [11] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. on Info. Theory*, vol. 35, no. 3, pp. 605–619, 1989.
- [12] R. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 773–785, 1993.
- [13] G. Kramer, "Directed information for channels with feedback," Ph.D. Dissertation, Swiss Federal Institute of Technology Zurich, 1998.
- [14] —, "Capacity results for the discrete memoryless network," *IEEE Trans. Inform. Theory*, vol. 49, pp. 4–21, 2003.
- [15] T. S. Han, *Information-Spectrum Method in Information Theory*. Springer, 2003.
- [16] —, "An information-spectrum approach to capacity theorems for the general multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2773–2795, 1998.
- [17] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [18] A. Lapidoth and I. Telatar, "The compound channel capacity of a class of finite-state channels," *IEEE Trans. Inform. Theory*, vol. 44, p. 973, 1998.
- [19] N. T. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inform. Theory*, vol. 21, pp. 100–102, 1975.
- [20] F. Alajaji, "Feedback does not increase the capacity of discrete channels with additive noise," *IEEE Trans. Inform. Theory*, vol. 41, pp. 546–549, 1995.
- [21] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Transactions on Information Theory*, vol. 31, no. 2, pp. 124–142, 1985.
- [22] Y. Kim, "A coding theorem for a class of stationary channels with feedback," Jan 2007, submitted to *IEEE Trans. Inform. Theory*. Available at [arxiv.org/cs.IT/0701041](http://arxiv.org/cs.IT/0701041).
- [23] J. Massey, "Causality, feedback and directed information," *Proc. Int. Symp. Information Theory Application (ISITA-90)*, pp. 303–305, 1990.
- [24] S. Tatikonda, "Control under communication constraints," Ph.D. dissertation, MIT, Cambridge, MA, 2000.
- [25] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback," *IEEE Trans. on Information theory*, vol. 51, pp. 780–789, 2005.
- [26] S. Yang, A. Kavcic, and S. Tatikonda, "Feedback capacity of finite-state machine channels," *IEEE Trans. Inform. Theory*, pp. 799–810, 2005.



- [27] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," Sep 2006, submitted to IEEE Trans. Inform. Theory. Available at [arxiv.org/pdf/cs.IT/0608070](http://arxiv.org/pdf/cs.IT/0608070).
- [28] H. H. Permuter, P. W. Cuff, B. Van-Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," Aug 2006, submitted to IEEE Trans. Inform. Theory. Available at [arxiv.org/cs.IT/0610047](http://arxiv.org/cs.IT/0610047).
- [29] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," September 2006, submitted to IEEE Trans. Inform. Theory. Available at [arxiv.org/cs.IT/0609139](http://arxiv.org/cs.IT/0609139).
- [30] B. Shrader and H. H. Permuter, "On the compound finite state channel with feedback," in *ISIT 2007*. Nice, France: IEEE, 2007.
- [31] G. Kramer, "Directed information for channels with feedback," Ph.D. Dissertation, Swiss Federal Institute of Technology Zurich, 1998.
- [32] S. Pradhan, "Source coding with feedforward: Gaussian sources," in *Proceedings 2004 International Symposium on Information Theory*, 2004, p. 212.
- [33] R. Venkataramanan and S. S. Pradhan, "Source coding with feedforward: Rate-distortion function for general sources," in *IEEE Information theory workshop (ITW)*, 2004.
- [34] R. Zamir, Y. Kochman, and U. Erez, "Achieving the gaussian rate-distortion function by prediction," July 2006, submitted for publication in "IEEE Trans. Inform. Theory". [Online]. Available: <http://www.eng.tau.ac.il/~zamir/papers/dpcm.pdf>
- [35] A. Rao, A. Hero, D. States, and J. Engel, "Inference of biologically relevant gene influence networks using the directed information criterion," in *ICASSP 2006 Proceedings*, 2006.
- [36] P. Mathai, N. C. Martins, and B. Shapiro, "On the detection of gene network interconnections using directed mutual information," in *ITA, San-Deigo*, 2007.
- [37] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channel," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1277–1290, 1989.
- [38] A. Goldsmith and P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. on Info. Theory*, vol. 42, pp. 868–886, 1996.
- [39] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948.
- [40] Y. Kim, "Feedback capacity of the first-order moving average gaussian channel," *IEEE Trans. Inform. Theory*, vol. 52, p. 3063, 2006.
- [41] —, "Feedback capacity of stationary Gaussian channels," 2006, submitted to IEEE Trans. Inform. Theory. Available at [arxiv.org/pdf/cs.IT/0602091](http://arxiv.org/pdf/cs.IT/0602091).
- [42] A. K. S. Yang and S. Tatikonda, "On the feedback capacity of power constrained Gaussian channels with memory," *submitted to IEEE Trans. Inform. Theory*, Oct. 2003.
- [43] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Information Theory*, vol. IT-2, pp. 8–19, 1956.
- [44] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Svst. Tech. J.*, vol. 42, pp. 1977–1997, 1963.
- [45] S. Diggavi and M. Grossglauser, "On information transmission over a finite buffer channel," *IEEE Trans. Inform. Theory*, vol. 52, p. 1226, 2006.
- [46] T. M. Cover, A. E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. 26, pp. 648–657, 1980.
- [47] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. 22, pp. 226–228, 1975.
- [48] G. Salinetti and R. Wets, "On the convergence of sequences of convex sets in finite dimensions," *SIAM Review*, vol. 21, no. 1, pp. 18–33, 1979.